

Supplementary Information for
Signal Peptides Generated by Attention-Based Neural Networks

Zachary Wu^{†,1}, Kevin K. Yang^{†,2}, Michael J. Liszka^{†,3}, Alycia Lee⁴, Alina Batzilla⁵, David Wernick³, David P. Weiner³, Frances H. Arnold^{*,1}

¹ Department of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA

² Generate Biomedicines, Cambridge, MA, USA; work performed while at California Institute of Technology

³ BASF Enzymes, San Diego, CA, USA

⁴ Department of Computational and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA

⁵ Department of Molecular Biotechnology, Heidelberg University, Heidelberg, Germany; work performed while at BASF Enzymes

*To whom correspondence may be addressed. Email: frances@cheme.caltech.edu

Tables

- a) p. 2 Table 1: Primers used to generate linear DNA fragments
- b) p. 2 Table 2: Enzyme activity reaction conditions
- c) p. 3 Table 3: *Bacillus* production strains
- d) p. 3 Table 4: Control sequences generated by the Transformer for random inputs
- e) p. 3 Table 5: Distribution of Protein and SP lengths as obtained from UniProt

Supplemental Sections

- I) p. 4 Comparing various generation approaches
- II) p. 7 Functionality classification from experimental validation
- III) p. 18 Activity assays at higher dilution
- IV) p. 25 Characteristics of functional vs nonfunctional generated SPs
- V) p. 26 All MSAs for functional SPs
- VI) p. 41 Bibliography

Supplemental Table 1: Primers used to generate linear DNA fragments

Primer	Sequence
pHT vector fwd	CAGACTTTCTAGAGGTCTCATAGCGCAGCC
pHT vector rev	TGAGCTCCTCGAGGGTCTCTATATAGAGTC
Signal Peptide fwd	TCCGCCTGACCTCCATGGGGTCTCAATATG
Signal peptide rev	CTCCACCTAGCCTGATATCGGTCTCTAGCC
Gene of Interest fwd	ATTCCGCCTGACCTGGTCTCAGGCTG
Gene of Interest rev	CTGAGCCTCCACCTAGCCTGGTCTCTGCTA

Supplemental Table 2: Enzyme activity reaction conditions

Enzyme Class	Substrate	Buffer	Reaction Conditions
Amylase	Red Starch (Megazyme)	50mM NaCl, 50mM BTP, 50mM Citric acid, 50mM N-cyclohexyl-3-aminopropanesulfonic acid (CAPS), pH 7.0	20ul of the cell supernatant were added to 10ul of 2% (w/v) Red Starch substrate in buffer and the mixture was incubated at 40 °C for 30 min. Reactions were quenched with 50ul of 95% ethanol and the assay plate was centrifuged at 4000 rpm for 10 min. The absorbance was measured in 60ul of the supernatant solution at 510nm.
Lipase	C8-PNP (4-Nitrophenyl octanoate, Sigma)	100mM HEPES, pH 7.5, 100mM NaCl, 20mM CaCl ₂ , 0.1% Triton-100	10ul of the cell supernatant were added to 90ul of 0.8 mM C8-PNP substrate in buffer. The absorbance was measured at 405nm kinetically for 15 min at a 31 sec interval. The initial reaction rate was calculated from the linear portion of the kinetic read.
Protease	AAPF (N-Succinyl-Ala-Ala-Pro-Phe p-nitroanilide, Sigma)	100mM HEPES buffer, pH 7.5	10ul of the cell supernatant were added to 90ul of 2mM AAPF substrate in buffer. The absorbance was measured at 410nm kinetically for 15 min at a 31 sec interval. The initial reaction rate was calculated from the linear portion of the kinetic read.
Xylanase	Azo-Xylan Birchwood (Megazyme)	50mM NaCl, 50mM Bis-tris propane (BTP), 50mM Citric acid, 50mM N-cyclohexyl-3-aminopropanesulfonic acid (CAPS), pH 7.0	20ul of the cell supernatant were added to 20ul of 2% (w/v) Azo-Xylan substrate in buffer and was incubated at 40 °C for 30 min. Reactions were quenched with 100ul of 95% Ethanol and the assay plate was centrifuged at 4000 rpm for 10 min. The absorbance was measured in 90ul of the supernatant solution at 590nm.

Supplemental Table 3: Enzyme activity reaction conditions

Strain	Modifications	Reference
<i>Bacillus subtilis</i> PY79 base strain	$\Delta nprE \Delta aprE \Delta epr \Delta mpr \Delta nprB \Delta vpr \Delta bpr \Delta sigF \Delta skfA \Delta xpf \Delta lytC \Delta sdpc$	Provided by BASF
BS – Prot	PY79 base strain $\Delta skfA \Delta spbC \Delta skfA$	Provided by BASF
BS – Amy	PY79 base strain $\Delta skfA \Delta spbC \Delta skfA \Delta AmyE$	Provided by BASF
BS – Xyl	PY79 base strain $\Delta skfA \Delta spbC \Delta skfA \Delta xynA \Delta xynC$	Provided by BASF
BS – Lip	PY79 base strain $\Delta skfA \Delta spbC \Delta skfA \Delta estA \Delta estB$	Provided by BASF

Supplemental Table 4: Control sequences generated by the trained Transformer for random inputs.

	Randomized Inputs	Generated SPs
Bacillus AA distribution	DTTRFTAIEIFTTHWSGMVLFMSIIIVVFFSIGDSGVGDANIFYHQARLFKA DKIANVQTLKGVYSNPLDYYFNHDLPRSGRLNLGNSHHIQLNKLQILF	MKLLVLLILILIPVVAWA
Bacterial AA distribution	SEVHPENTYQEGLPGIKALRIQFMTATIVHEGSTIDEVVRKKALTQTTICIPY LIAASIIVSSGEDMHIIIVIKHMAQQVLPKLISCAMEHFVGPShAEASE	MNWLIVIMFAIGLACSTSLA
Uniform AA distribution	DIVPSLYPPAKSDEKVELTSKQSELNSENKATQQILKDGYPATPGRVRKID AETKKQLNLIKHTYRTVRKQEDKGVYQASNPRNFKFIVPQLVQYKGRN	MEHTLRILVCSLLFTVAIWG

Supplemental Table 5: Distribution of Protein and SP lengths, as obtained from UniProt.

Length (Amino Acids)	Signal Peptide	Protein
Mean \pm StDev	23.5 \pm 6.2	422.5 \pm 454.1
Minimum	10	21
Median	22	323
Maximum	69	13076

Supp Section I: Comparing various generation approaches

In this section, we compare four methods of generating SP sequences for their efficacy: (A) a profile Hidden Markov Model (pHMM), (B) a heuristic-based generation approach, (C) a variational autoencoder (VAE)¹ trained on just SP sequences, and (D) the Transformer-based approach. In the first round of comparison, we query the SignalP 5.0 server with these SP sequences prepended to protein sequences to obtain SignalP's probability of functioning.

- (A) pHMM: We first generated an alignment with Clustal Omega² of SPs from our training. We used 4000 randomly selected SPs, as Clustal Omega is limited to 4000 sequences. From this point on, we use default settings in S. Eddy's HMMER package.³ From the Clustal Omega alignment, we then use HMMER's *hmmbuild* to build a profile HMM. From this profile HMM, we use HMMER's *hmmemit* to generate protein samples. As HMMs typically follow the first order Markov property, the HMMs do not have a concept of length, and can generate longer sequences. We generate 1024 constructs by emitting 1024 SP sequences from *hmmemit*, each of which is prepended to one of the 41 enzymes tested.
- (B) Heuristics: We follow the following heuristics from a recent review by Owji, Hajar, et al.⁴ The N-region (1-6 residues) contains at least 1 positive residue (KR). The H-region (7-15 residues) contains hydrophobic residues (AILMFYWV). The C Region (3-7 residues) contains uncharged residues (STNQAVILMFYW) followed by an AXA motif, where X is not in C/P/A. 768 SPs were generated by heuristics and randomly matched to one of the 41 input proteins.
- (C) VAE: We train a VAE with fully connected layers and ReLU activation on SP sequences. The sizes of the hidden layers are 400 and 20 each for the mean and variance of the latent encoding. We then randomly sample 1024 SP sequences, each of which is prepended to one of the 41 enzymes tested.
- (D) Transformer: As described in the main text.

We then query these constructs with SignalP v5.0 to obtain the probability of containing a Signal Peptide in gram-positive hosts.⁵ The results are shown visually in the figure below, and summarized in its accompanying table.

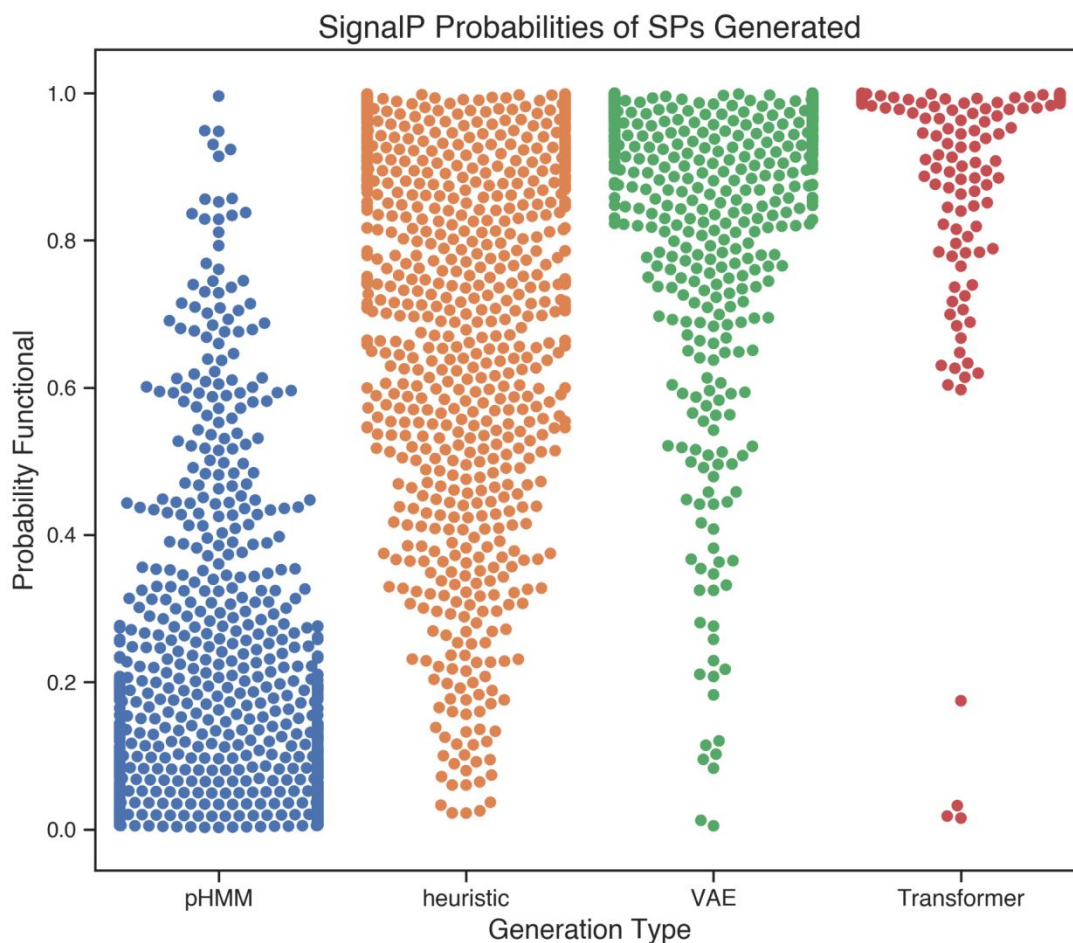


Figure S1: Probability of generated sequences as predicted by SignalP 5.0 The sequences are generated by either a pHMM, heuristic-based approach, VAE, or Transformer model.

Table S1: Summary of probability of predictions for various generation methods, as predicted by SignalP 5.0

Method	Mean SignalP Probability
pHMM	15.6% \pm 19.2%
heuristic	70.8% \pm 25.3%
VAE	92.4% \pm 15.2%
Transformer	90.4% \pm 17.1%

The VAE and Transformer significantly outperform the other two methods (p -value $< 10^{-20}$), but are not statistically different from each other (p -value = 0.251). As both are predicted to function with high probability by SignalP, we turn to other methods of analysis below.

First, we compare the log-likelihood and percent (unaligned) sequence identity on an identical, withheld validation set of 5707 SP sequences for the VAE and Transformer, which are comparable between the two methods, although the VAE is more accurate by this metric.

Method	Log likelihood on Validation Set	% Sequence Identity (Unaligned)
VAE	-2.86	45.4%
Transformer	-3.01	34.5%

However, in closer examination of the sequences generated by the VAE and Transformer, an issue of repeated residues begins to emerge:

Sample VAE Sequences	Sample Transformer Sequences
MRKRLALALAALSLLLLLSFGVKALAGSGA	MKFTQAVLSLLGSAATALA
MRLLLLLLLVLLLAAPPPGLS	MKLKKGVLIAICLGISSTFA
MKLLLLLVTLTSTVLALQA	MRVLSATAFLALLAFGLSGATA
MRLLLLALLAAAAVALASA	MLFKSVLLALASAGVAVNA
MASSSSSLFVVLAVLLLLLLLLTLSSA	MNISIFVGKLALAAALGSALVA

Namely, the hydrophobic region appears to be particularly saturated in repeating residues for VAE sequences. For a more comprehensive comparison, we determined the average length of substrings with maximum R repeating residues (e.g.: an example of $R=1$ is LLLLLLL, $R=2$ is LLLILIL, $R=3$ is LALLAI).

	Average longest substring from VAE Seqs	Average longest substring from Transformer Seqs
R = 1	3.9 \pm 1.9	2.1 \pm 0.6
R = 2	7.2 \pm 3.0	4.0 \pm 1.0
R = 3	10.7 \pm 3.5	6.3 \pm 1.5

As biological sequences rarely contain such long repeats, we elect to proceed with the Transformer generated sequences in experimental validation.

Supp Section II: Functionality classification from experimental validation

Classification Explanation

SP-enzyme constructs were generated stochastically and tested in biological replicates. Thus, some random SP-enzyme constructs contain the same SP and enzyme but are not part of the same set of biological replicates.

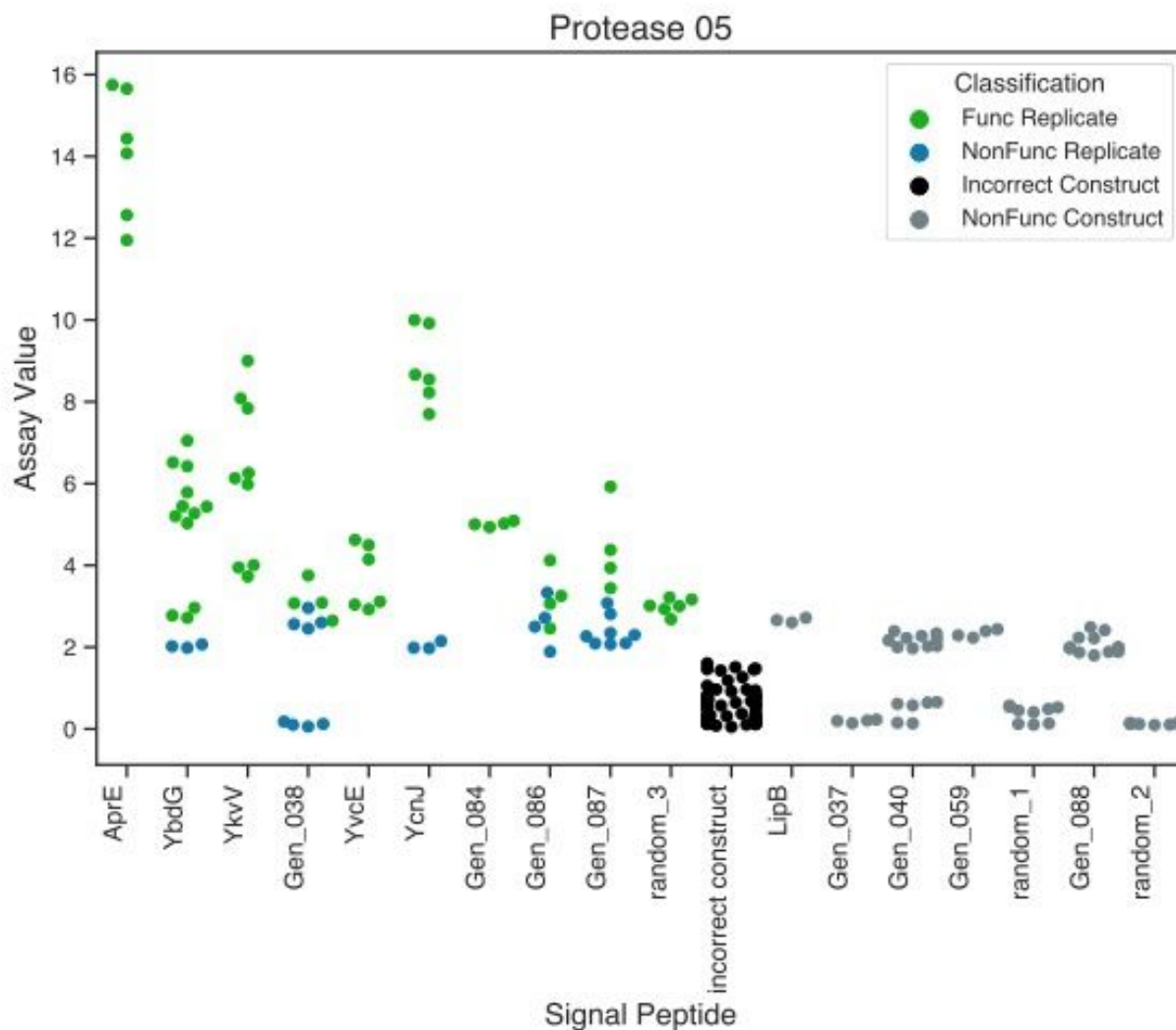
An SP-enzyme construct is classified as functional if any set of biological replicates satisfies two conditions (p-value and effect size):

- 1) The p-value for the set of replicates compared to the negative controls (constructs without either an SP or enzyme) was less than 0.05 in a two-sided t-test with unequal variance for two independent samples of scores. Biological replicates are grown from the same colony, after which they are processed separately. The null hypothesis is that these two independent samples have identical expected values. The `scipy.stats` implementation (`ttest_ind`) was used. (p-value)
- 2) The difference between the mean of the replicates and the mean of the incorrect constructs is greater than 2 times the difference between the mean of the incorrect constructs and the highest measured activity of the incorrect constructs. (effect size)

The second condition is incorporated to gain some concept of effect size: the *activity* of the SP-enzyme construct must be high, not just statistically significant.

In the following plots, constructs classified as functional are to the left of the “incorrect constructs”, which are shown in black. For functional constructs, some sets of replicates (picked from the same colony) would have been classified as nonfunctional. These are shown in blue. The functional sets of replicates are shown in green. Finally, constructs classified as nonfunctional are shown in grey. Summary statistics and notes highlighting important points are also provided.

All generated Signal Peptides can be found in Supplemental File 1.



Shuffled_ID: seq1
 Original_ID: Arnold_036
 Plot_ID: Protease 05

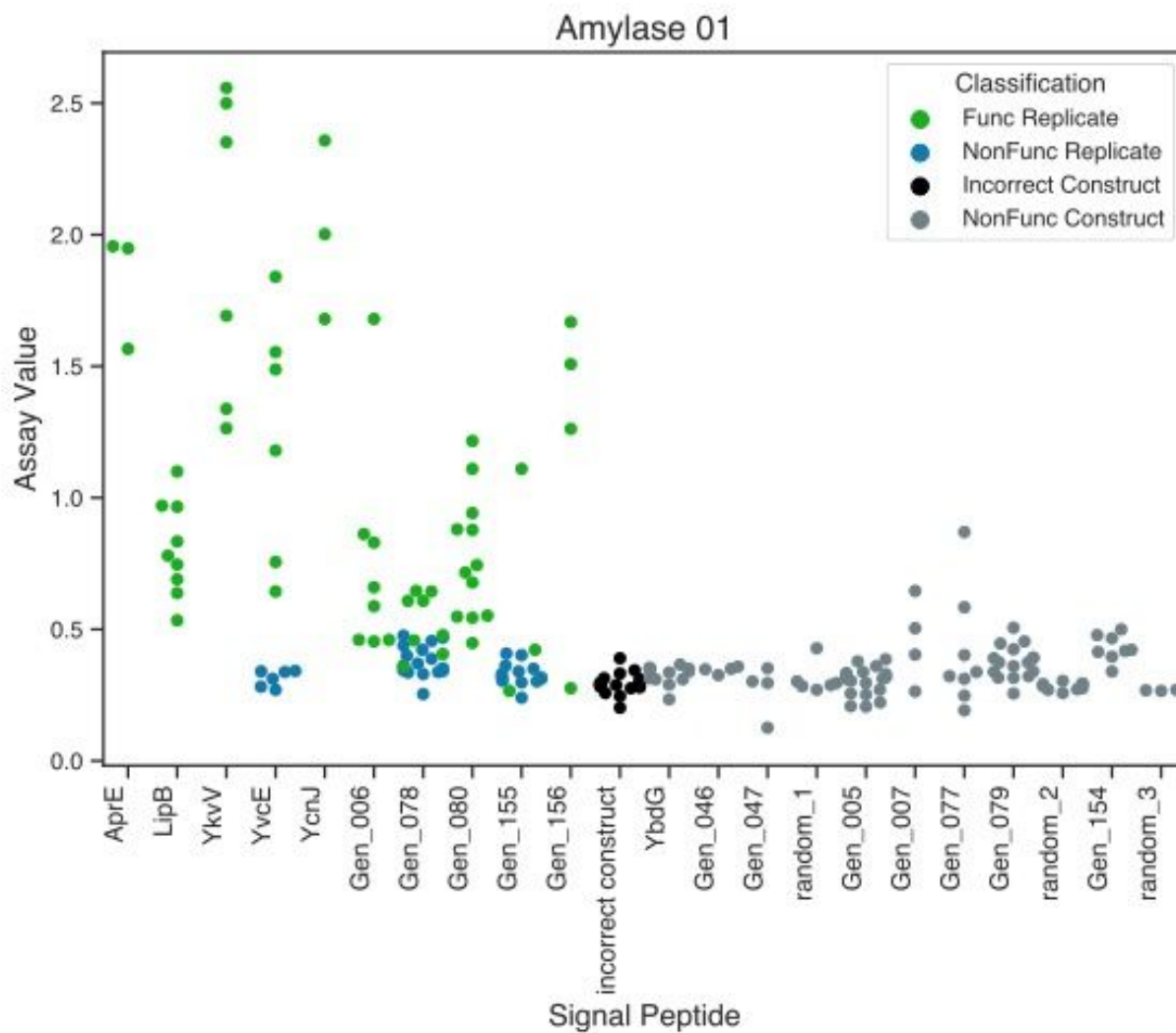
Construct Functionality Classification Summary

0.33 : 1 out of 3 constructs with SPs generated for random inputs functional*
 0.83 : 5 out of 6 constructs with natural SPs functional
 0.50 : 4 out of 8 constructs with generated SPs functional

Assay information is provided in Supplemental Table 2.

*The SP sequence for the control generated for a randomized input that was classified as functional is

MEHTLRILVCSLLFTVAIWG.

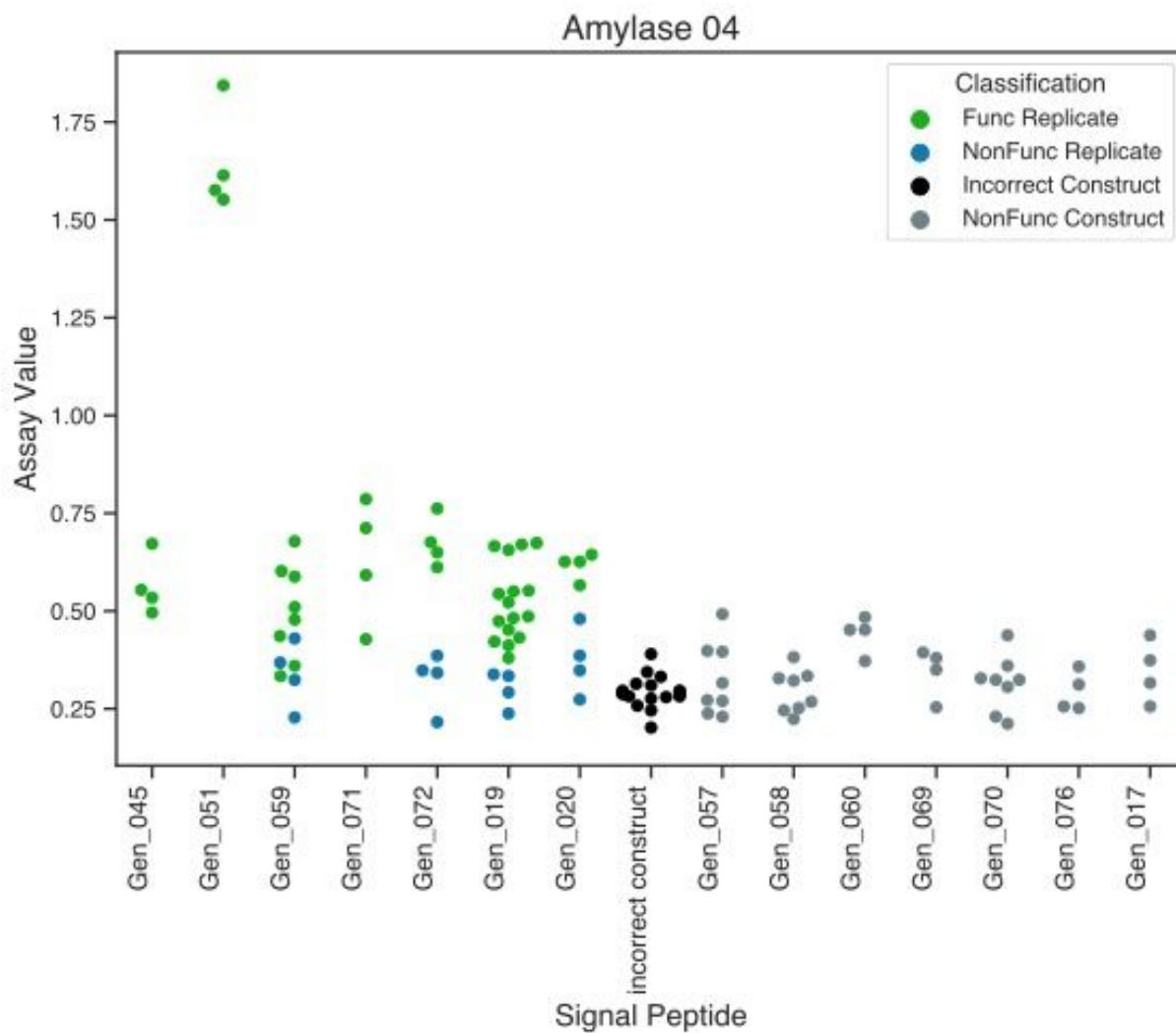


Shuffled_ID: seq39
 Original_ID: Arnold_001
 Plot_ID: Amylase 01

Construct Functionality Classification Summary

0.00 : 0 out of 3 constructs with SPs generated for random inputs functional
 0.83 : 5 out of 6 constructs with natural SPs functional
 0.42 : 5 out of 12 constructs with generated SPs functional

Assay information is provided in Supplemental Table 2.



Shuffled_ID: seq5
 Original_ID: Arnold_004
 Plot_ID: Amylase 04

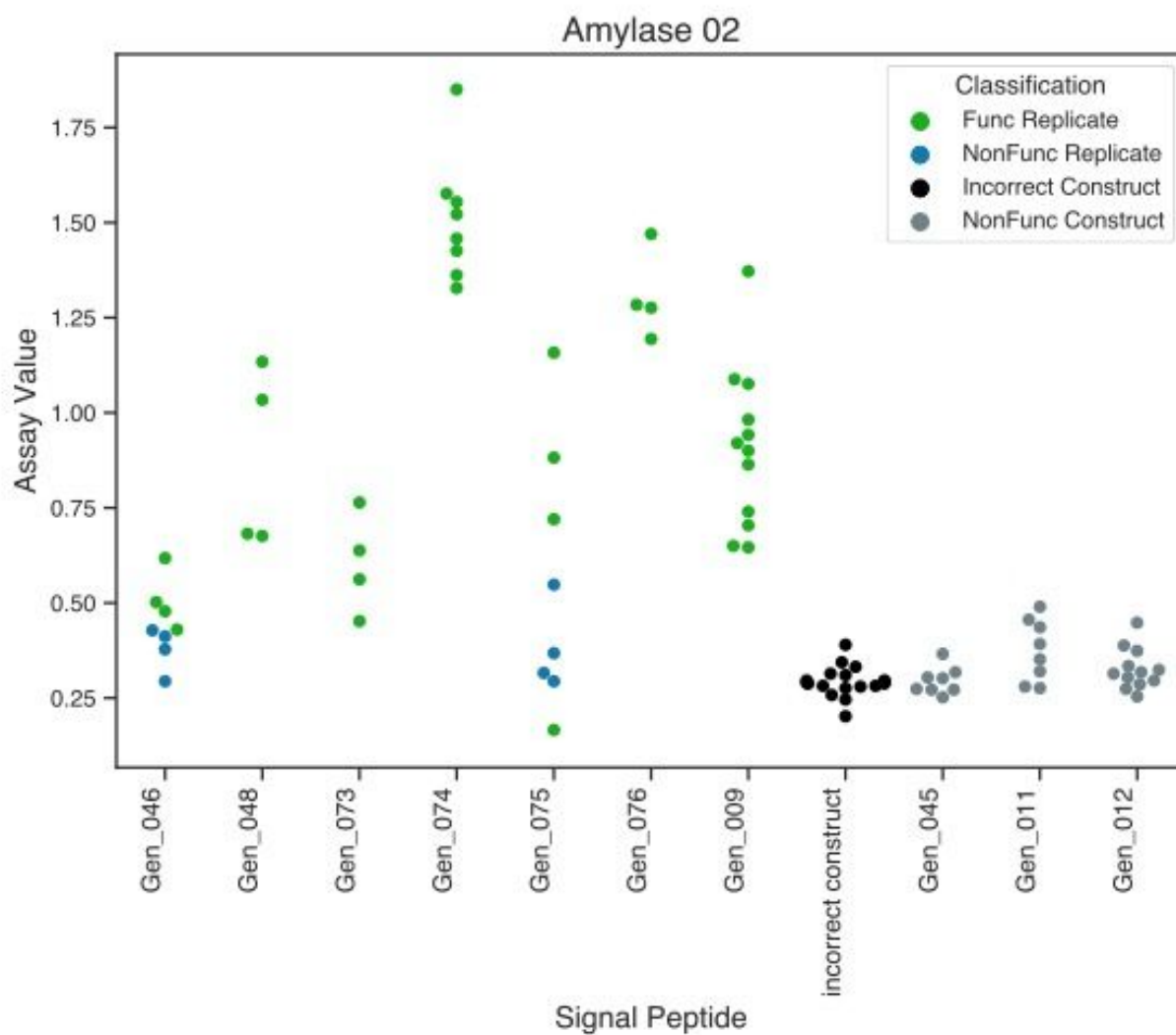
Construct Functionality Classification Summary

No constructs with SPs generated for random inputs tested.

No constructs with natural SPs tested.

0.50 : 7 out of 14 constructs with generated SPs functional

Assay information is provided in Supplemental Table 2.



Shuffled_ID: seq19
 Original_ID: Arnold_002
 Plot_ID: Amylase 02

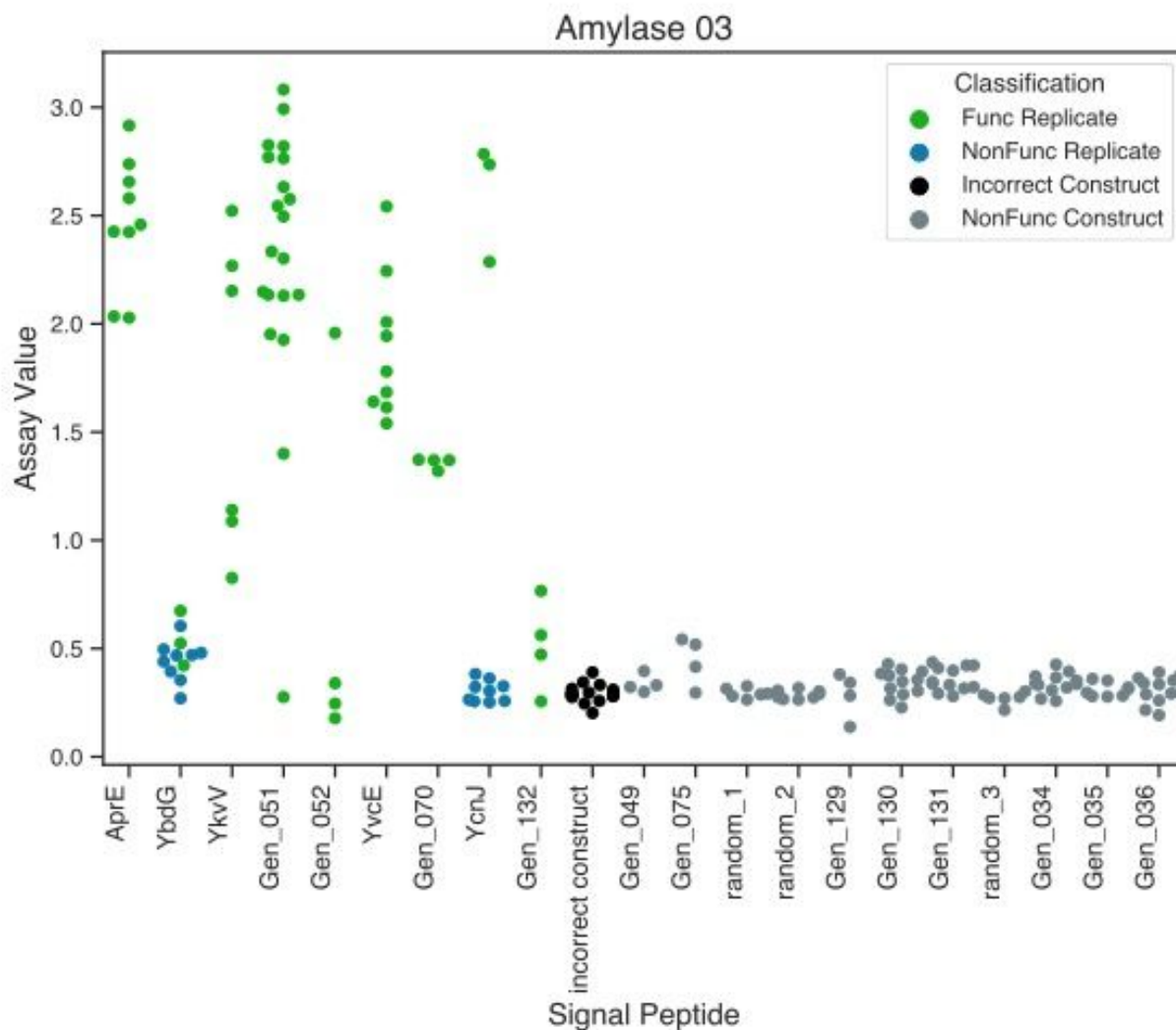
Construct Functionality Classification Summary

No constructs with SPs generated for random inputs tested.

No constructs with natural SPs tested.

0.70 : 7 out of 10 constructs with generated SPs functional

Assay information is provided in Supplemental Table 2.

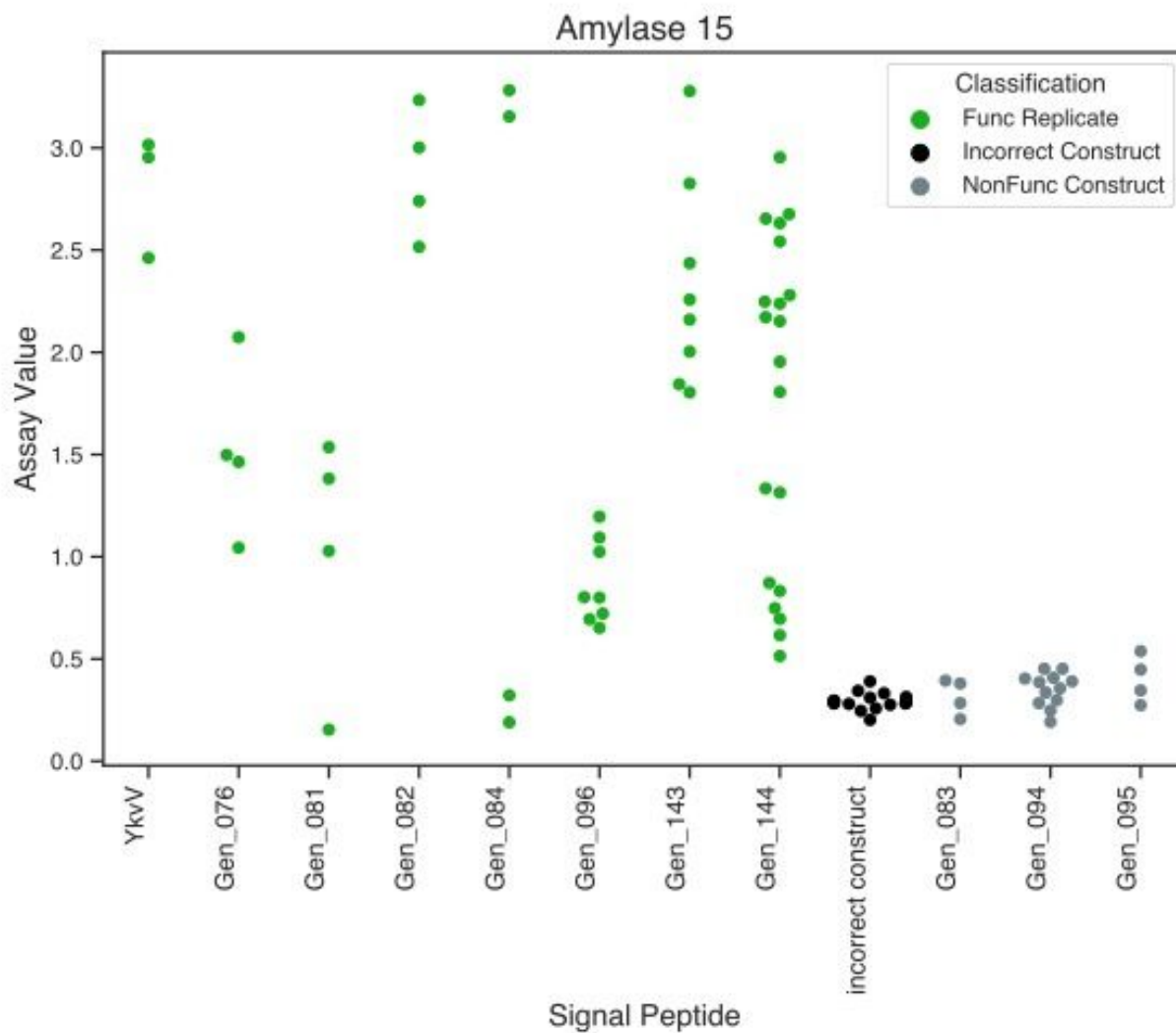


Shuffled_ID: seq33
 Original_ID: Arnold_003
 Plot_ID: Amylase 03

Construct Functionality Classification Summary

0.00 : 0 out of 3 constructs with SPs generated for random inputs functional
 1.00 : 5 out of 5 constructs with natural SPs functional
 0.33 : 4 out of 12 constructs with generated SPs functional

Assay information is provided in Supplemental Table 2.



Shuffled_ID: seq21
 Original_ID: Arnold_015
 Plot_ID: Amylase 15

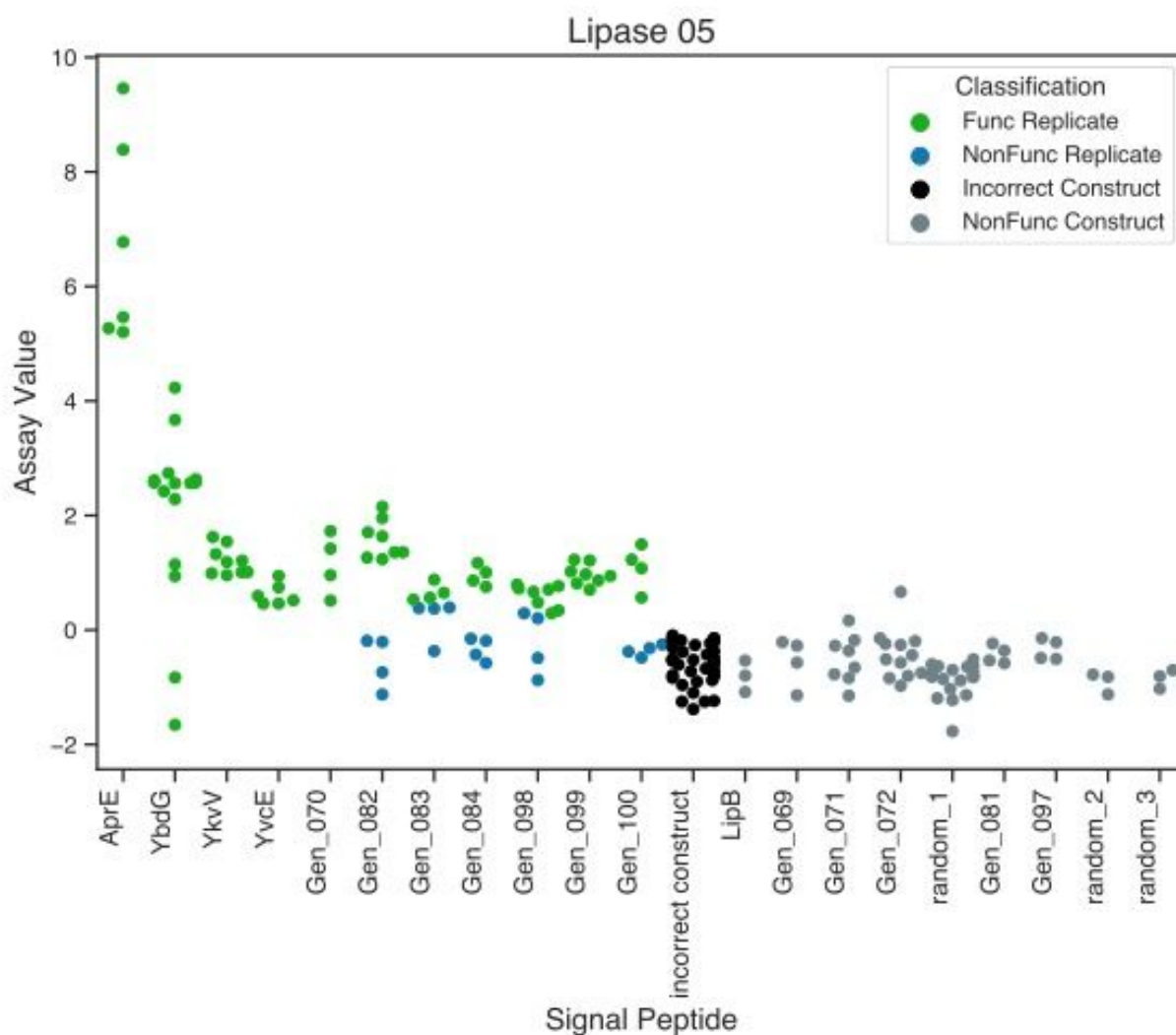
Construct Functionality Classification Summary

No constructs with SPs generated for random inputs tested.

1.00 : 1 out of 1 constructs with natural SPs functional

0.70 : 7 out of 10 constructs with generated SPs functional

Assay information is provided in Supplemental Table 2.

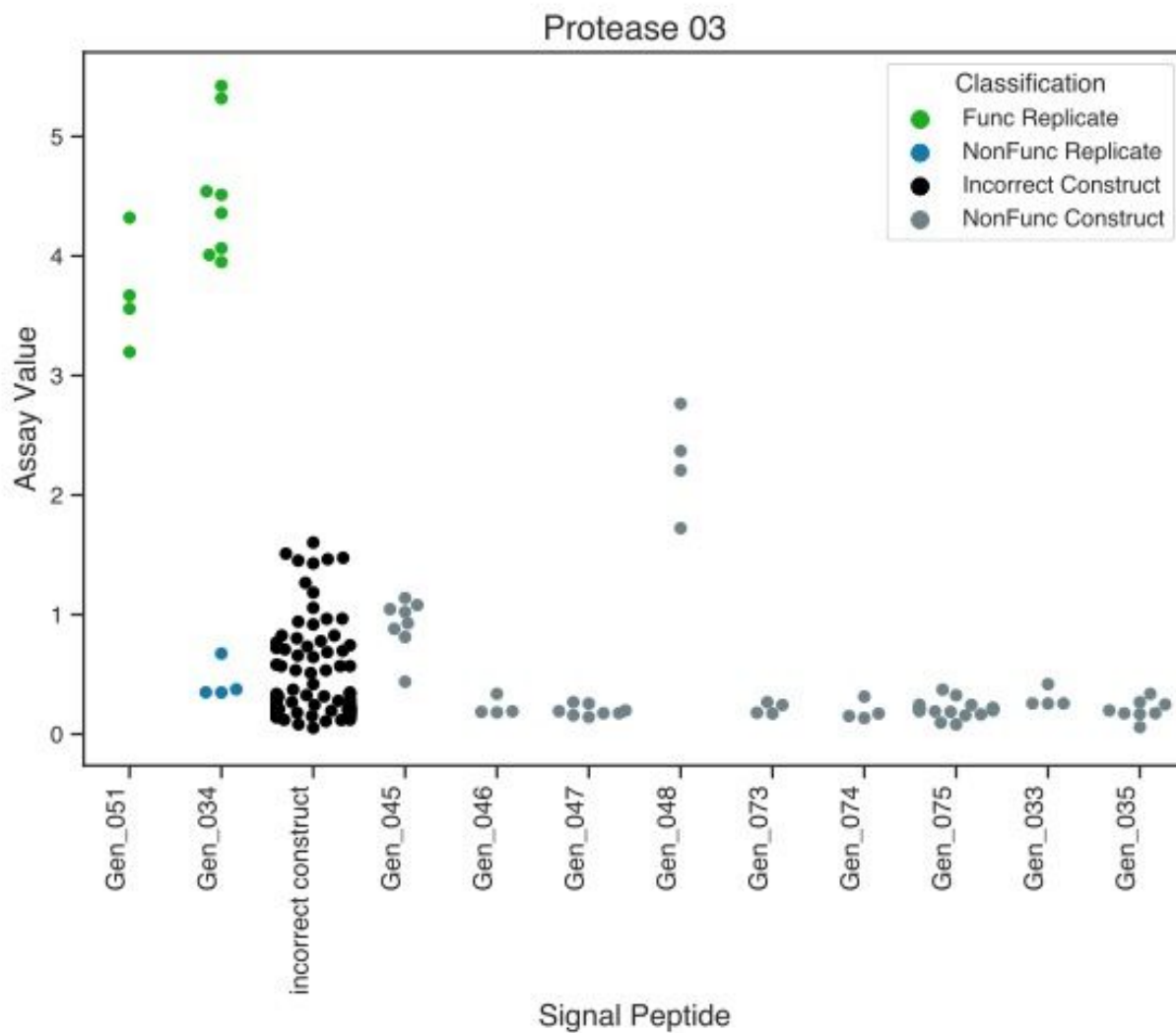


Shuffled_ID: seq25
 Original_ID: Arnold_026
 Plot_ID: Lipase 05

Construct Functionality Classification Summary

0.00 : 0 out of 3 constructs with SPs generated for random inputs functional
 0.80 : 4 out of 5 constructs with natural SPs functional
 0.58 : 7 out of 12 constructs with generated SPs functional

Assay information is provided in Supplemental Table 2.



Shuffled_ID: seq36
 Original_ID: Arnold_034
 Plot_ID: Protease 03

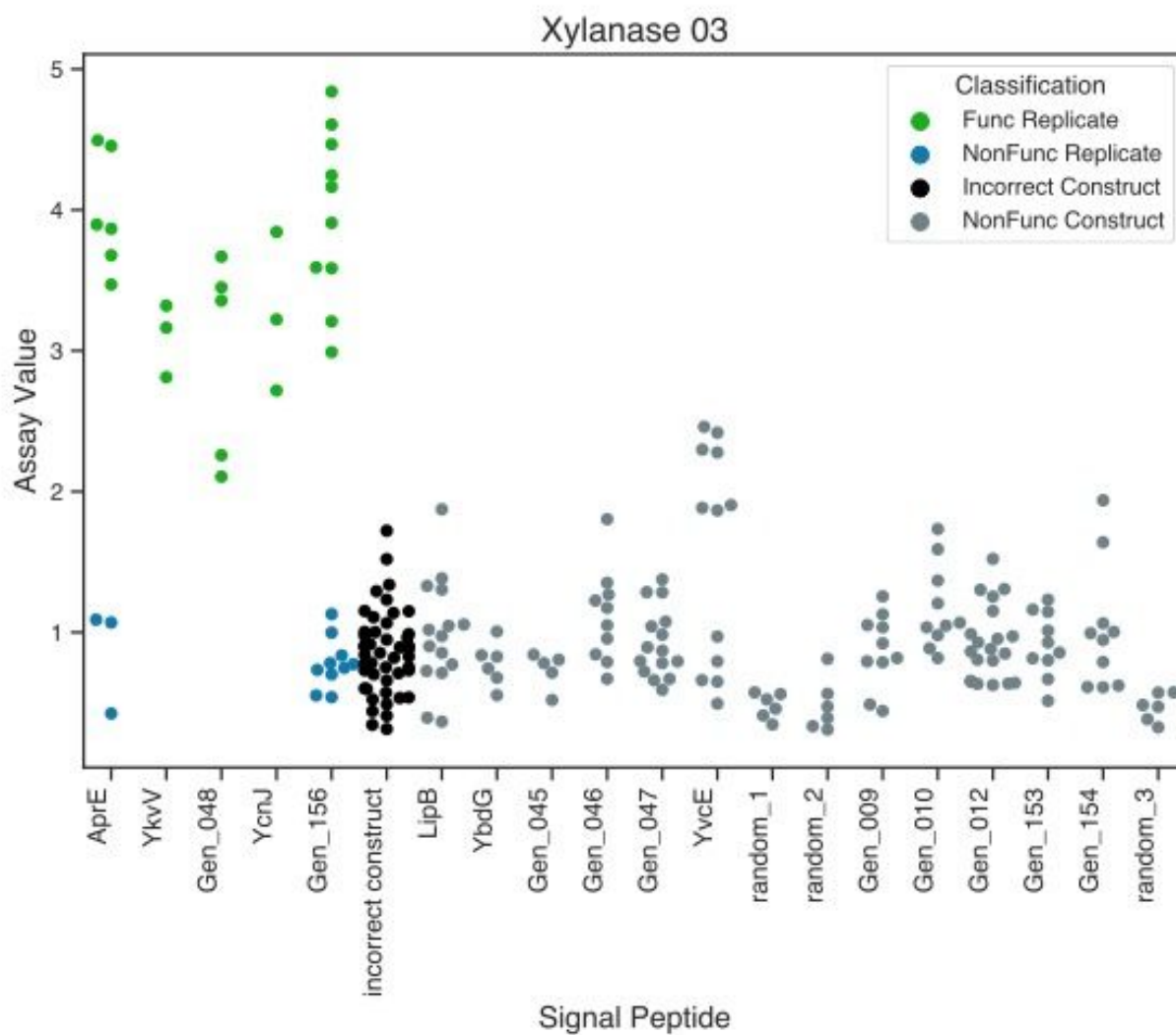
Construct Functionality Classification Summary

No constructs with SPs generated for random inputs tested.

No constructs with natural SPs tested.

0.18 : 2 out of 11 constructs with generated SPs functional

Assay information is provided in Supplemental Table 2.

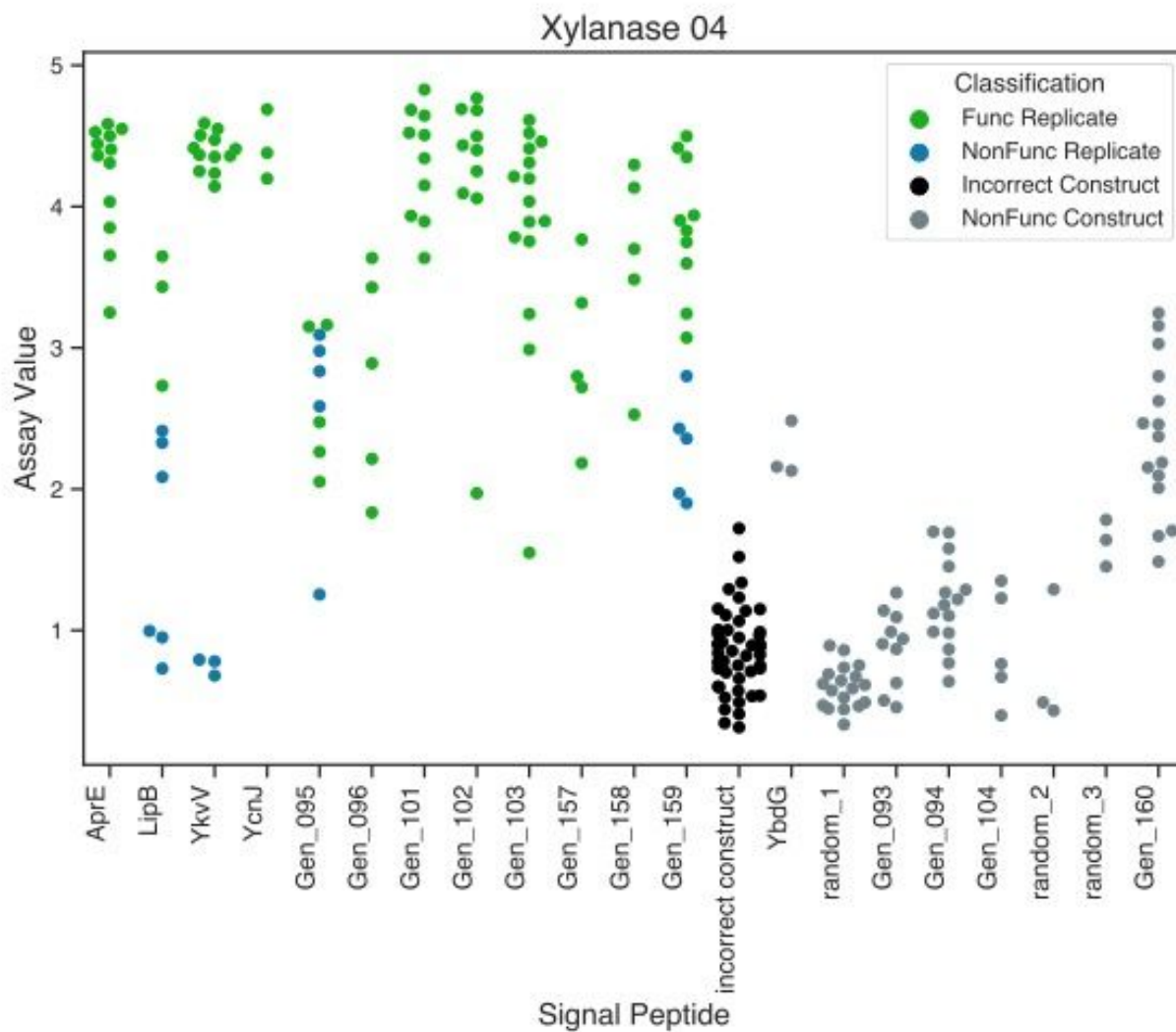


Shuffled_ID: seq3
 Original_ID: Arnold_039
 Plot_ID: Xylanase 03

Construct Functionality Classification Summary

0.00 : 0 out of 3 constructs with SPs generated for random inputs functional
 0.50 : 3 out of 6 constructs with natural SPs functional
 0.20 : 2 out of 10 constructs with generated SPs functional

Assay information is provided in Supplemental Table 2.



Shuffled_ID: seq26
 Original_ID: Arnold_040
 Plot_ID: Xylanase 04

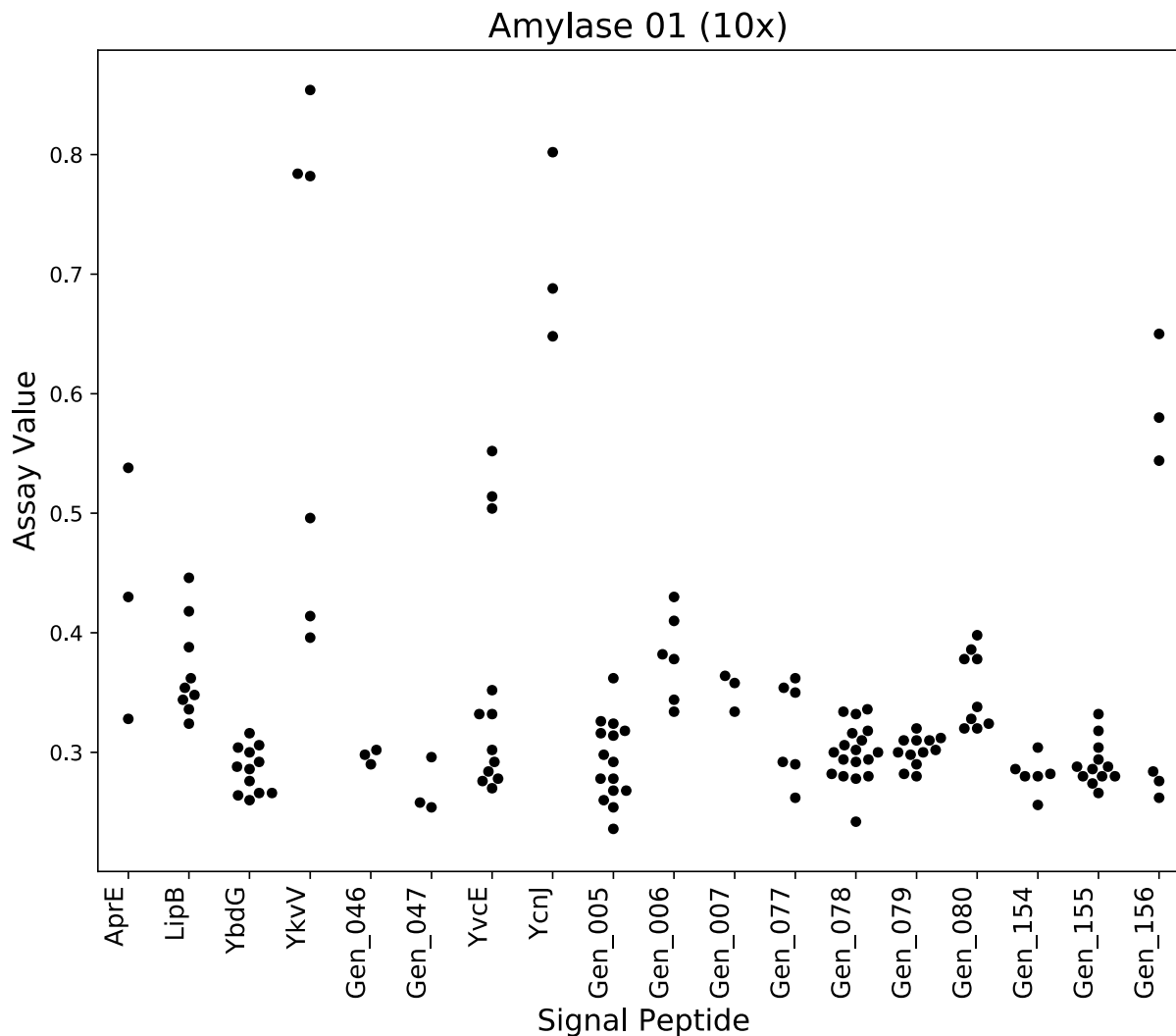
Construct Functionality Classification Summary

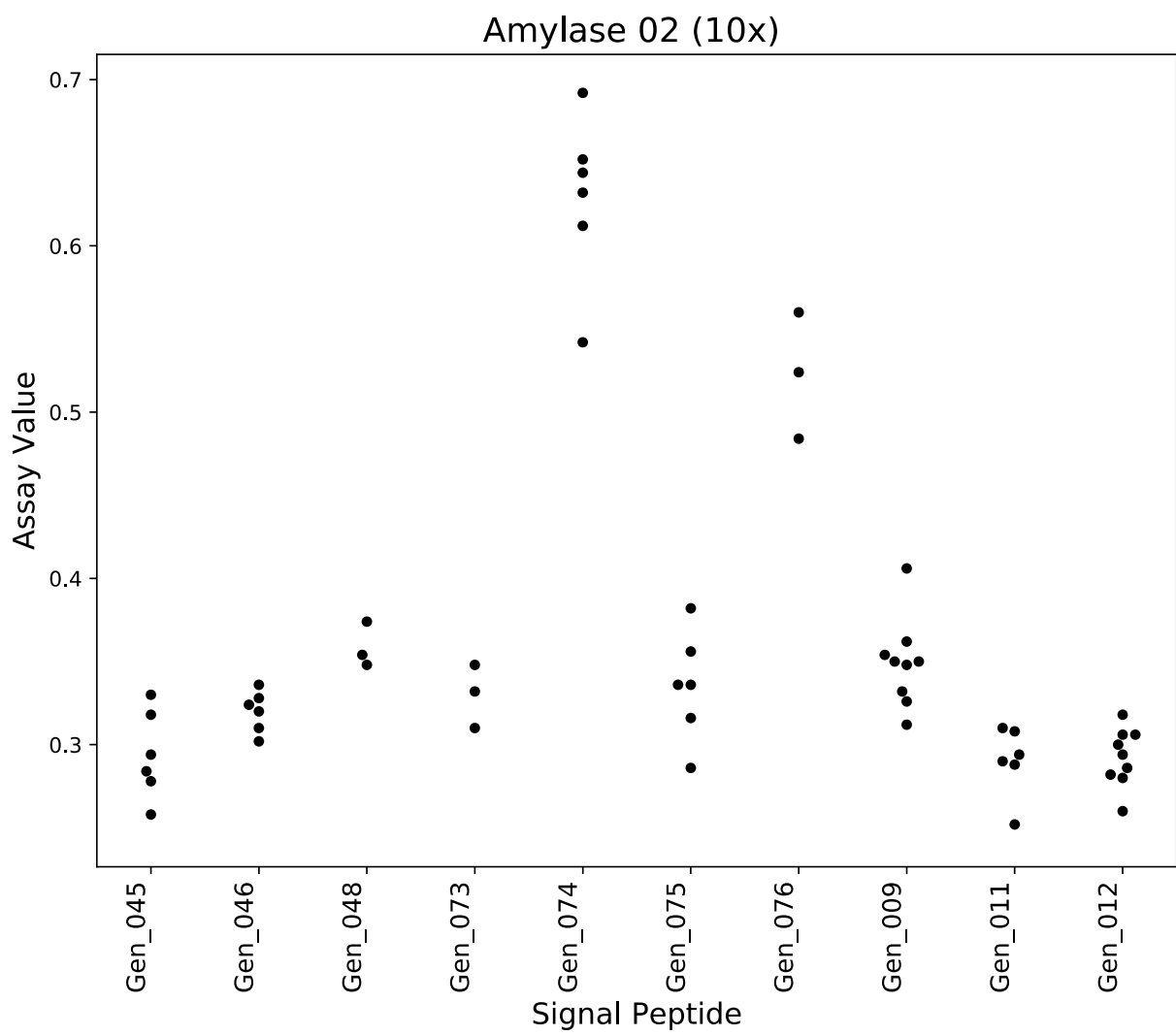
0.00 : 0 out of 3 constructs with SPs generated for random inputs functional
 0.80 : 4 out of 5 constructs with natural SPs functional
 0.67 : 8 out of 12 constructs with generated SPs functional

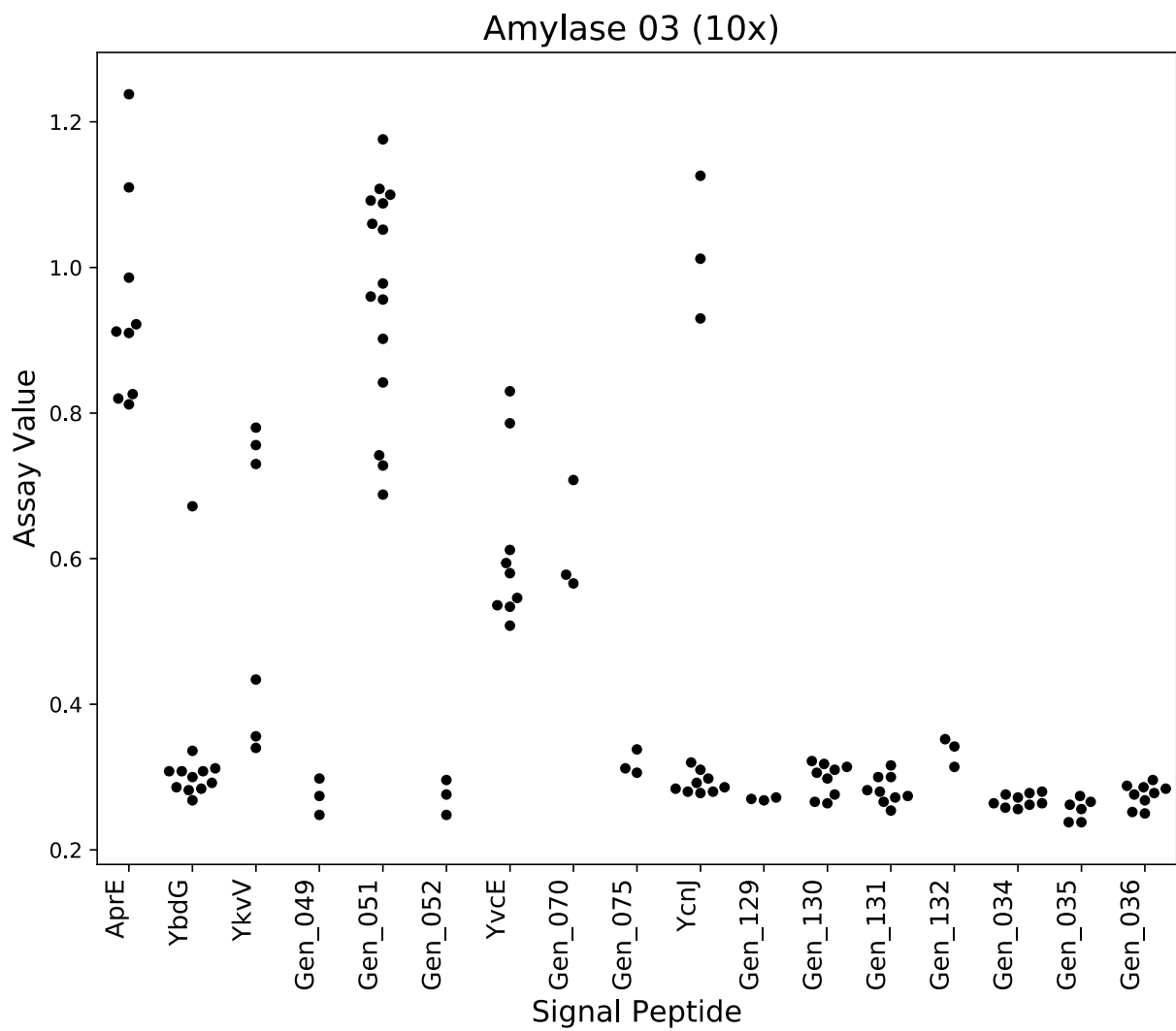
Assay information is provided in Supplemental Table 2.

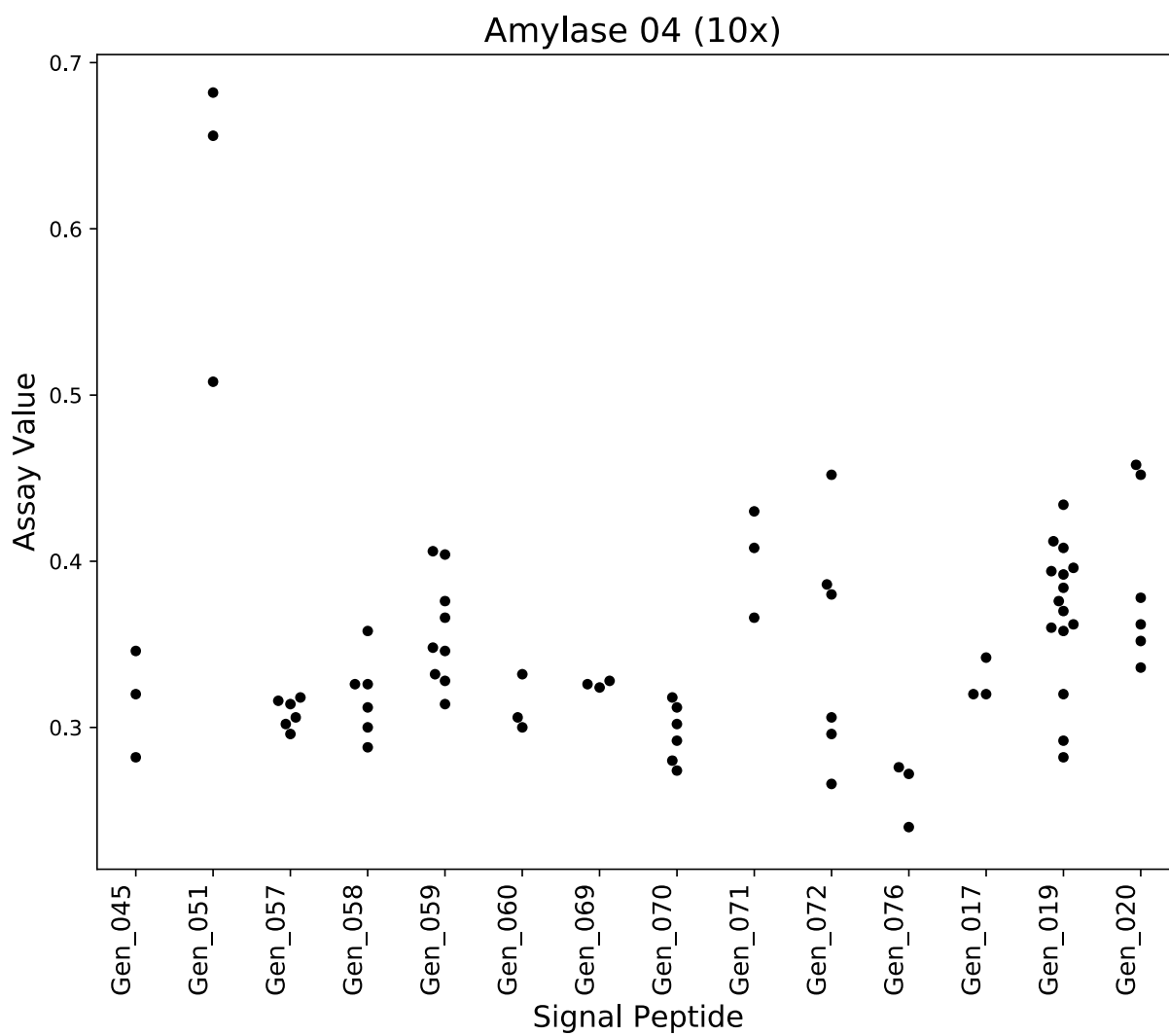
Supp Section III: Activity assays at higher dilution

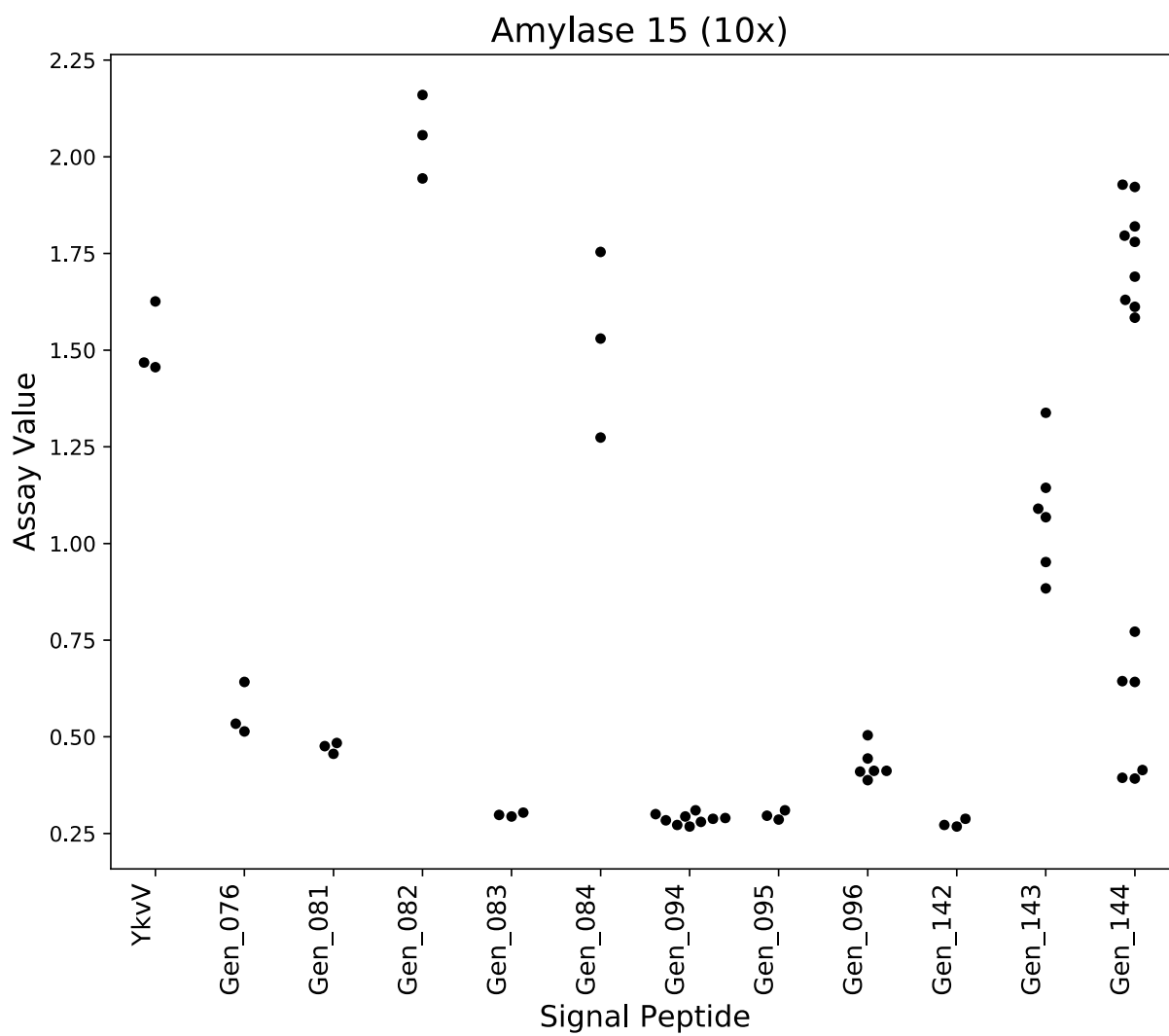
While Supplemental Section 2 shows activity data for identifying functional constructs, we found that some amylase and xylanase constructs were outside of the linear range (activity was too high) to make accurate comparisons between the high functioning constructs. To address this, we performed the assays at higher dilution (10x for amylase, 50x for xylanase) to obtain data in the linear range of the assay.

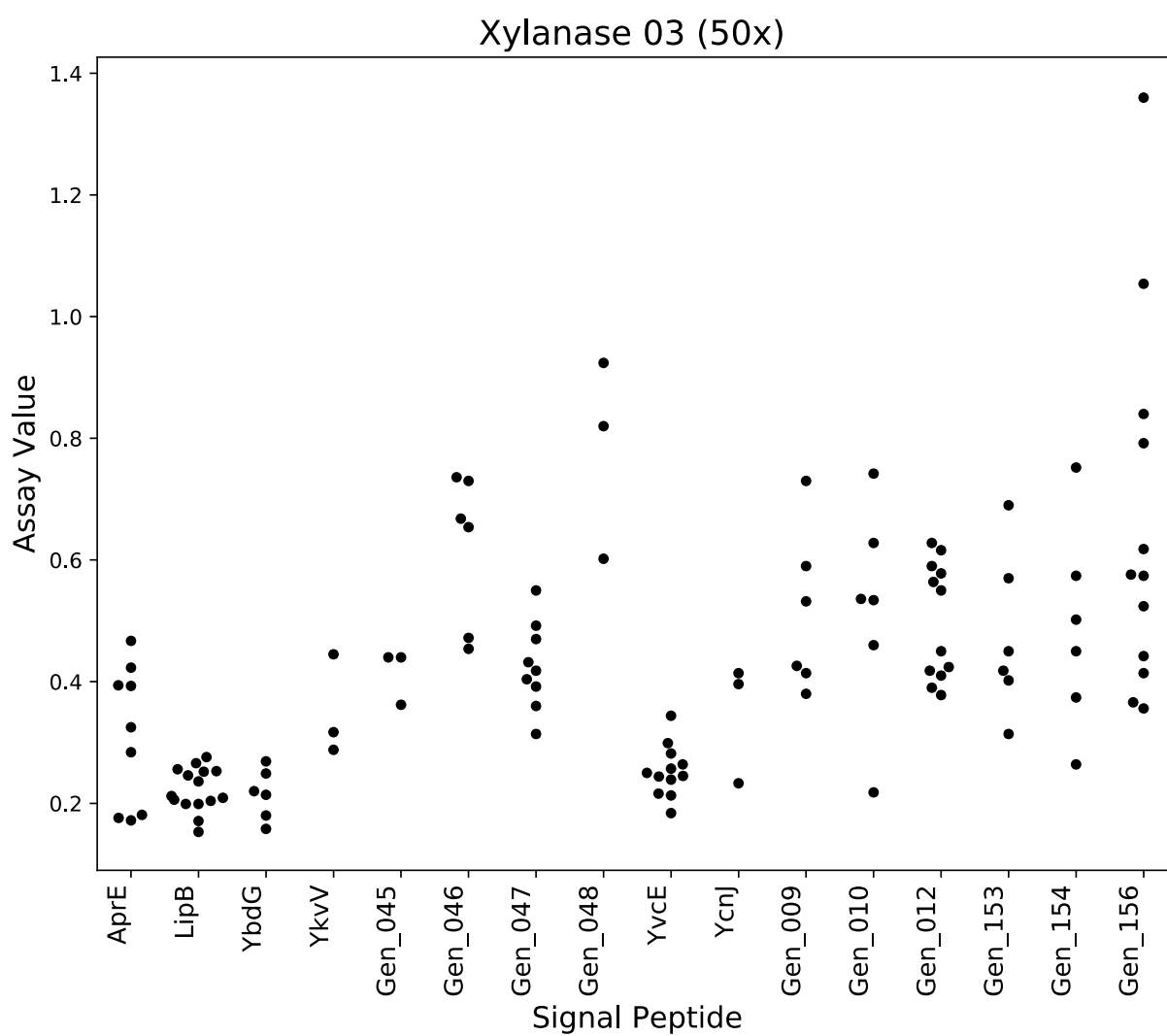


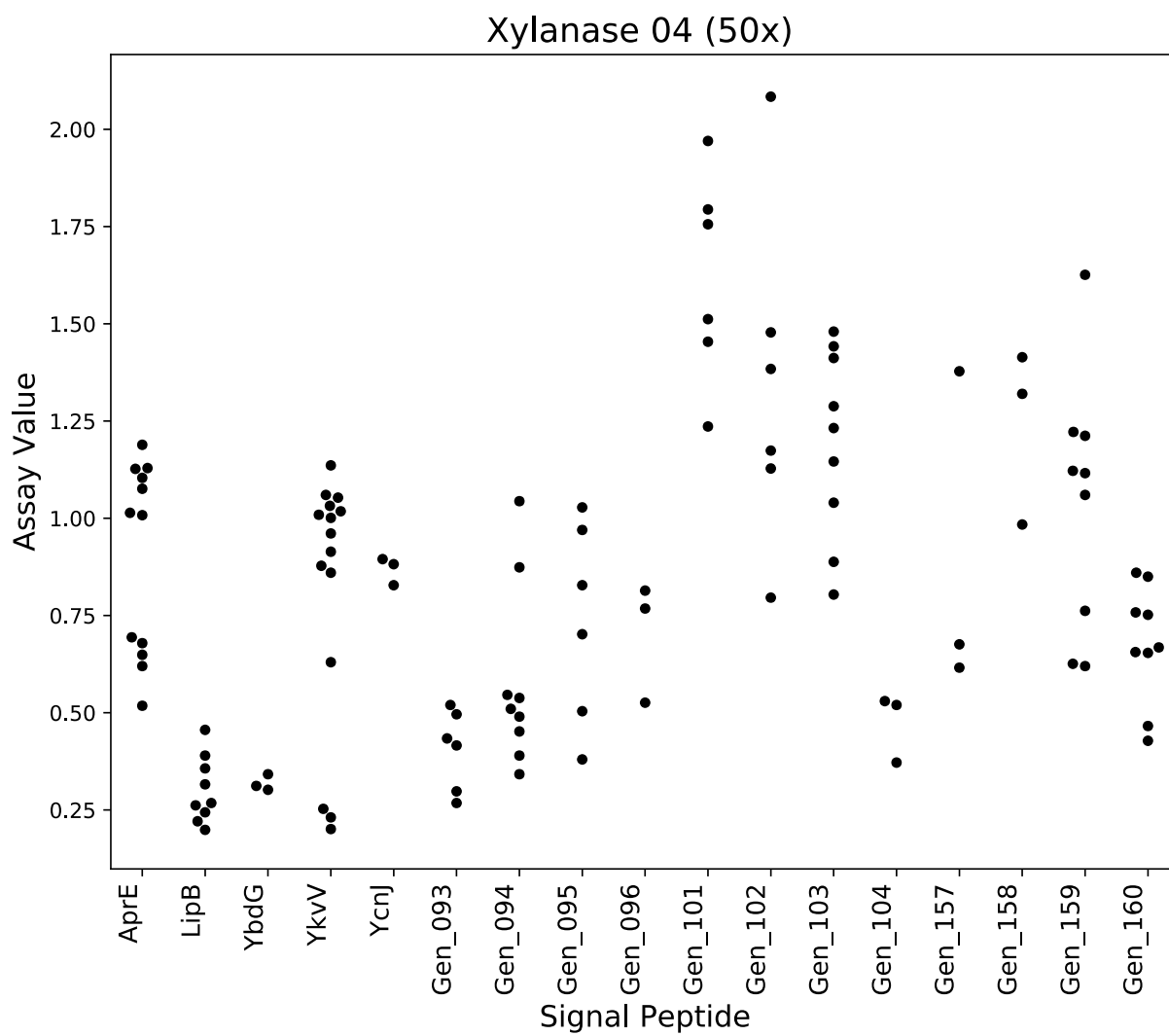












Supp Section IV: Characteristics of functional vs nonfunctional generated SPs

We attempted to identify general characteristics of functional versus nonfunctional signal peptides that may help in discerning whether a signal peptide is functional or not. To do so, we separated the 27 signal peptides that are only present in functional constructs, and 30 signal peptides that are only present in nonfunctional constructs, and attempted to identify a distinguishing property. No property was identified for distinguishing between functional and nonfunctional constructs (p-value < 0.05) and SignalP 5.0 was also not able to discern between functional and nonfunctional generated SPs (ROC=0.59).

Protein Property	p-value
Length	0.9229
Hydrophobicity Index	0.3870
Hydrophilicity Index	0.2126
Molecular Weight (MW)	0.7963
Average MW	0.1170
Aromaticity	0.9935
Instability Index	0.9748
Isoelectric Point	0.3243

Supp Section V: All MSAs for functional SPs

MSAs for the top 10 closest matching natural Signal Peptides (by sequence identity) to each of the 43 generated SPs that were functional in any construct are provided. Alignments are sorted by closest sequence identity to the generated natural Signal Peptide (shown first).

All generated Signal Peptides can be found in Supplemental File 1.

Sequence: MGFRLKALLVGCLIFLAVSSAIA

gen_sp	M	G	F	R	L	-	-	-	-	-	K	A	L	L	V	G	C	L	I	F	L	A	V	S	S	A	I	A	-
nat_sp0	M	K	R	L	-	-	-	-	-	-	S	A	L	L	L	T	C	L	L	S	A	V	S	S	L	S	A	L	A
nat_sp1	M	G	L	K	V	S	S	-	-	-	S	L	L	C	L	T	I	L	L	A	V	S	S	I	V	S	A	-	-
nat_sp2	M	L	I	R	K	W	K	-	-	-	A	G	L	L	A	G	L	S	I	L	A	L	A	S	S	A	D	A	-
nat_sp3	M	R	V	L	P	-	-	-	-	-	L	A	L	L	V	G	L	L	A	V	S	D	A	-	-	-	-	-	-
nat_sp4	M	K	Y	R	I	L	M	A	T	L	L	A	V	C	L	G	I	F	S	L	S	A	P	A	F	A	-	-	-
nat_sp5	M	R	F	R	P	S	I	V	A	L	L	S	V	C	F	G	L	L	T	F	L	Y	S	G	S	A	F	A	-
nat_sp6	M	G	F	T	K	I	L	V	-	-	T	F	F	L	V	G	L	L	V	I	S	S	S	P	Q	N	A	I	A
nat_sp7	M	G	R	L	-	-	-	-	-	-	L	A	L	V	V	G	A	A	L	V	S	S	A	C	-	-	-	-	-
nat_sp8	M	K	K	L	-	-	-	-	-	-	T	L	V	L	F	G	M	L	F	L	A	S	S	A	H	A	-	-	-
nat_sp9	M	G	R	L	C	T	K	-	-	-	F	L	T	S	V	G	C	L	I	L	L	L	V	T	G	S	G	S	-
					5					10					15					20								25	

Sequence: MIKTLLVSSILIPCLATGA

gen_sp	-	-	M	I	K	T	L	L	V	S	S	I	L	I	P	C	L	A	T	G	A	-	-	-	-	-	-	-	-
nat_sp0	-	M	I	I	K	T	L	L	A	S	V	A	I	M	L	I	A	T	V	N	A	-	-	-	-	-	-	-	-
nat_sp1	-	M	K	R	N	T	L	L	A	L	V	L	V	I	L	I	F	P	T	L	S	T	A	-	-	-	-	-	-
nat_sp2	-	-	-	M	K	K	L	L	L	A	A	S	I	I	C	L	A	S	A	G	L	A	-	-	-	-	-	-	-
nat_sp3	-	-	-	M	K	K	T	L	L	G	S	L	I	L	L	A	F	A	T	N	A	-	-	-	-	-	-	-	-
nat_sp4	-	-	M	I	A	T	L	L	S	S	L	L	L	T	G	P	I	S	A	G	A	-	-	-	-	-	-	-	-
nat_sp5	-	-	-	M	K	K	L	L	L	T	A	S	I	I	C	L	A	S	A	G	L	A	-	-	-	-	-	-	-
nat_sp6	M	N	I	K	L	T	L	L	V	L	I	S	I	I	N	L	M	I	I	Q	P	I	Q	T	L	A	-	-	-
nat_sp7	M	V	F	M	N	K	T	L	L	V	S	S	L	I	A	C	L	S	I	A	S	V	N	V	Y	A	-	-	-
nat_sp8	-	-	-	M	K	K	T	L	L	A	S	S	L	I	A	C	L	S	I	A	S	V	N	V	Y	A	-	-	-
nat_sp9	-	M	R	V	S	T	L	V	L	S	T	S	I	I	P	I	A	T	A	-	-	-	-	-	-	-	-	-	-
					5					10					15					20								25	

Sequence: MIRLKRLLAGLLLPLFVTAFG

gen_sp	-	M	I	R	L	K	R	L	L	A	G	L	L	L	P	L	-	F	V	T	A	F	G
nat_sp0	M	I	L	R	R	L	L	L	A	G	S	L	L	L	A	-	S	F	A	T	A	-	-
nat_sp1	M	I	L	R	R	L	L	L	A	G	S	L	L	L	A	T	A	F	T	S	A	-	-
nat_sp2	M	I	L	R	R	L	V	L	A	G	S	L	L	L	A	T	A	F	T	S	A	-	-
nat_sp3	-	M	L	R	K	L	T	P	L	A	L	A	L	L	P	L	V	A	G	-	-	-	-
nat_sp4	-	-	-	-	M	K	L	L	K	S	S	L	L	L	L	L	P	F	V	T	A	-	-
nat_sp5	-	-	M	H	G	L	L	L	A	G	L	L	A	L	P	L	N	V	F	A	-	-	-
nat_sp6	-	M	C	R	L	R	V	L	-	-	L	L	L	L	P	L	A	F	V	S	S	S	A
nat_sp7	M	I	L	R	R	L	L	L	A	G	S	L	L	L	A	T	S	F	T	S	A	-	-
nat_sp8	-	-	M	H	G	L	L	L	A	G	L	L	A	L	P	L	N	V	L	A	-	-	-
nat_sp9	-	M	L	G	R	S	L	L	A	L	L	P	F	V	G	L	A	F	S	-	-	-	-

Sequence: MKCCRIMFVLLGLWFVFGLSVPGGRTEA

gen_sp	M	K	-	-	C	C	R	I	M	F	V	L	L	G	L	W	F	V	F	G	L	S	V	P	G	G	R	T	E	A
nat_sp0	M	K	-	-	C	C	R	I	M	F	V	L	L	G	L	W	F	V	F	G	L	S	V	P	G	G	R	T	E	A
nat_sp1	-	-	-	-	M	G	R	F	I	F	V	S	F	G	L	L	V	V	F	-	L	S	L	S	G	-	-	T	E	A
nat_sp2	-	-	-	-	M	G	R	F	I	F	V	S	F	G	L	L	V	V	F	-	I	S	L	S	G	-	-	T	E	A
nat_sp3	-	-	-	-	M	G	R	F	I	F	V	S	F	G	L	L	V	M	F	-	L	S	L	S	G	-	-	T	E	A
nat_sp4	M	K	-	-	K	V	R	F	I	F	L	-	-	A	L	L	V	F	L	-	A	S	P	E	G	-	-	A	M	A
nat_sp5	M	K	R	V	R	V	K	V	I	F	V	S	F	G	L	L	V	V	F	-	L	S	L	S	G	-	-	T	A	A
nat_sp6	-	-	-	-	M	G	R	I	I	F	V	S	F	G	L	L	V	V	F	-	L	S	L	S	G	-	-	T	G	A
nat_sp7	-	-	-	-	M	G	R	F	I	F	V	S	F	G	L	L	V	V	F	-	L	S	L	R	G	-	-	T	G	A
nat_sp8	-	-	-	-	M	G	R	F	I	F	V	S	F	G	L	L	V	V	F	-	L	S	L	S	G	-	-	S	E	A
nat_sp9	M	K	-	-	K	A	A	A	V	L	L	S	L	G	L	V	F	G	F	-	S	Y	G	A	G	H	V	A	E	A

Sequence: MKFFNPFKVIALACISGALATAQA

gen_sp	M	K	F	F	N	P	F	-	-	-	K	V	I	A	L	A	C	I	S	G	A	L	A	T	A	Q	A
nat_sp0	M	R	A	F	K	-	-	-	-	-	W	A	L	A	I	G	A	T	L	A	L	P	L	T	A	Q	A
nat_sp1	M	K	S	F	-	-	-	-	-	-	T	V	I	A	L	A	A	V	A	L	L	A	T	L	G	Q	A
nat_sp2	-	-	M	F	T	R	K	I	Q	K	T	A	L	A	M	L	I	S	G	A	M	A	G	T	A	Y	A
nat_sp3	M	K	N	K	K	R	V	L	I	A	S	S	L	S	C	A	I	L	L	L	S	A	A	T	T	Q	A
nat_sp4	M	K	F	L	-	-	-	-	-	-	-	T	A	L	S	A	I	G	A	L	V	A	T	A	T	A	A
nat_sp5	M	Q	F	K	-	-	-	-	-	-	N	V	A	L	A	A	S	V	A	A	L	S	A	T	A	S	A
nat_sp6	M	K	N	W	-	-	-	-	-	-	I	K	V	A	V	A	A	I	A	L	S	A	A	T	V	Q	A
nat_sp7	M	S	F	S	N	Y	K	V	I	A	M	P	V	L	V	A	N	F	V	L	G	A	A	T	A	W	A
nat_sp8	M	K	F	F	A	Y	F	A	V	-	I	A	L	S	S	A	S	L	I	N	L	F	K	R	A	T	A
nat_sp9	M	K	F	S	-	-	-	-	-	-	-	-	-	L	S	V	A	L	S	L	A	A	A	T	A	Q	A

Sequence: MKFLILATLSIFTGILA

gen_sp	-	-	-	-	M	K	F	L	I	L	-	A	T	L	S	-	-	I	F	T	G	I	-	L	A
nat_sp0	-	-	-	-	M	K	F	Q	L	L	-	-	-	-	T	L	V	S	I	A	T	T	T	L	A
nat_sp1	-	-	-	-	M	K	F	L	V	I	-	L	T	L	C	-	-	-	I	A	G	A	-	I	A
nat_sp2	-	-	-	-	M	K	F	L	N	I	-	L	T	L	A	-	-	F	I	T	G	M	A	S	A
nat_sp3	M	W	L	K	I	Q	V	F	L	L	A	I	T	L	I	T	L	G	I	Q	A	-	-	-	-
nat_sp4	M	K	L	L	S	K	T	F	L	I	-	L	T	L	T	F	F	F	F	G	I	A	-	L	A
nat_sp5	-	-	-	-	M	A	V	F	L	L	-	-	A	T	S	T	I	M	F	P	T	K	I	E	A
nat_sp6	-	-	-	-	M	K	F	L	I	L	-	T	A	L	C	A	V	T	L	A	-	-	-	-	-
nat_sp7	-	-	-	-	M	K	V	L	T	A	-	I	A	L	S	A	I	A	F	T	G	A	-	V	A
nat_sp8	-	-	-	-	M	K	T	F	L	I	-	L	A	L	L	A	I	V	A	T	T	A	R	I	A
nat_sp9	-	-	-	-	M	R	F	L	I	L	F	L	T	L	S	-	-	-	L	G	G	I	-	D	A
					5					10					15						20				25

Sequence: MKFLSTAFVLLIALVAGCSTA

gen_sp	M	K	F	L	S	T	A	F	V	L	L	I	A	L	V	A	G	C	S	T	A
nat_sp0	M	K	F	L	S	T	A	A	A	L	L	V	C	L	A	P	V	S	T	T	A
nat_sp1	M	K	F	L	I	A	F	A	V	L	A	L	V	A	C	I	N	A	-	-	-
nat_sp2	M	K	F	L	S	L	A	F	V	L	G	L	L	A	L	A	N	A	-	-	-
nat_sp3	M	K	F	L	T	A	L	S	A	I	G	A	L	V	A	T	A	T	A	-	-
nat_sp4	M	K	I	L	L	F	V	T	L	I	A	L	A	F	V	A	L	C	S	A	-
nat_sp5	M	K	F	S	L	V	A	T	I	V	L	L	A	L	A	Q	G	S	F	A	-
nat_sp6	M	K	R	L	F	L	S	F	V	A	-	L	A	L	L	A	G	S	I	A	A
nat_sp7	M	K	F	S	L	I	A	A	V	A	L	L	A	L	A	Q	G	S	F	A	-
nat_sp8	M	K	L	T	L	V	I	L	A	L	V	A	C	V	T	A	-	-	-	-	-
nat_sp9	M	K	L	S	T	V	L	L	S	A	G	L	A	S	T	T	L	A	-	-	-
					5					10					15					20	

Sequence: MKFQDLTLVLSLSTALA

gen_sp	M	K	F	Q	D	L	T	-	-	-	-	L	V	L	S	L	S	T	A	L	A
nat_sp0	-	-	M	K	Y	L	T	-	-	-	-	L	L	T	V	L	S	T	A	L	A
nat_sp1	-	-	M	K	F	L	T	-	-	-	-	P	L	V	L	S	S	L	A	S	A
nat_sp2	M	K	F	Q	L	L	T	-	-	-	-	L	V	S	I	A	T	T	T	L	A
nat_sp3	-	-	M	K	L	F	Q	I	F	P	L	L	L	S	L	T	S	V	T	L	A
nat_sp4	-	-	M	K	F	L	S	-	S	L	V	V	L	G	L	S	A	Q	A	L	A
nat_sp5	M	K	I	K	L	L	T	-	L	A	V	A	S	L	V	S	V	N	A	L	A
nat_sp6	-	-	M	K	F	L	T	-	-	-	-	-	L	A	L	S	A	T	A	T	A
nat_sp7	-	-	M	R	S	L	T	-	L	L	L	S	L	S	T	A	L	R	S	V	A
nat_sp8	-	-	M	K	F	L	Q	I	I	P	V	L	L	S	L	T	S	T	T	L	A
nat_sp9	M	K	F	R	L	T	A	L	A	V	A	A	L	L	T	S	T	A	S	F	A
					5					10					15					20	

Sequence: MKKKIAITLLFLSLLNRA

gen_sp	M	K	K	K	I	-	-	-	A	I	T	L	L	F	L	S	L	L	N	R	A	-	-	-
nat_sp0	-	M	K	K	I	-	-	V	A	I	F	L	V	F	L	G	S	L	W	A	-	-	-	-
nat_sp1	M	K	K	Q	I	-	-	F	F	T	L	I	L	L	I	S	G	L	A	R	A	-	-	-
nat_sp2	M	K	K	K	L	-	-	K	L	T	S	L	L	G	L	S	L	L	I	M	T	A	-	-
nat_sp3	M	K	K	T	L	-	-	I	I	L	T	V	L	L	L	S	V	L	T	A	A	-	-	-
nat_sp4	M	K	K	K	I	P	S	L	A	L	G	I	L	L	V	F	L	L	Q	Y	L	V	A	G
nat_sp5	M	K	K	A	K	A	I	F	L	F	I	L	I	V	S	G	F	L	L	V	A	-	-	-
nat_sp6	M	K	K	T	T	-	-	L	K	F	A	A	L	T	L	L	G	L	S	N	L	A	L	A
nat_sp7	M	K	R	K	I	-	-	I	A	I	S	L	F	L	Y	I	P	L	S	N	A	-	-	-
nat_sp8	M	K	K	N	I	-	-	F	I	T	S	L	L	I	L	L	L	L	S	S	-	-	-	-
nat_sp9	M	K	K	N	I	-	-	I	A	G	C	L	F	S	L	F	S	L	S	A	L	A	-	-

Sequence: MKKKIVAVLTLSVVLA

gen_sp	-	M	K	K	K	I	V	A	V	L	T	L	S	V	V	L	A	-	-	-	-
nat_sp0	M	K	K	K	I	V	A	G	A	V	T	L	L	S	V	A	V	L	A	A	-
nat_sp1	M	K	K	K	I	F	A	G	A	V	T	L	L	S	V	A	V	L	A	A	-
nat_sp2	M	K	K	I	I	V	V	S	L	F	A	L	V	V	V	L	A	G	-	-	-
nat_sp3	M	K	L	K	L	I	V	L	A	L	T	M	S	V	V	T	A	Q	A	-	-
nat_sp4	M	K	N	K	F	A	A	L	V	I	T	L	F	S	V	L	A	-	-	-	-
nat_sp5	-	M	K	L	I	I	L	V	A	L	T	L	A	A	V	V	A	-	-	-	-
nat_sp6	M	V	K	K	T	I	A	A	I	F	S	V	L	V	L	S	S	V	L	T	A
nat_sp7	-	-	M	K	I	V	A	L	T	L	V	A	F	V	A	L	A	G	A	-	-
nat_sp8	M	V	K	K	T	I	A	A	I	F	S	V	L	V	L	S	T	V	L	T	A
nat_sp9	M	K	S	S	K	I	V	A	I	L	L	A	S	L	F	S	G	S	V	L	A

Sequence: MKKLLILACLLISSLES

gen_sp	-	M	K	K	L	L	I	L	A	C	L	L	I	S	S	L	E	S	-	-	-	-
nat_sp0	M	K	L	K	S	L	L	I	A	C	L	L	S	S	L	S	F	S	A	L	A	-
nat_sp1	-	M	K	W	I	T	L	I	C	L	L	I	S	S	T	L	I	E	S	-	-	-
nat_sp2	M	K	K	L	L	F	I	T	A	P	L	L	L	S	V	L	T	A	S	-	-	-
nat_sp3	-	M	K	W	I	T	L	I	C	L	L	I	S	S	S	F	I	E	S	-	-	-
nat_sp4	M	K	L	W	A	I	L	A	V	C	I	L	L	L	S	S	V	S	S	-	-	-
nat_sp5	M	K	N	K	S	K	L	L	A	C	C	L	M	A	L	P	I	S	S	F	S	-
nat_sp6	M	K	L	K	L	L	L	I	P	L	L	G	S	S	L	L	L	S	A	-	-	-
nat_sp7	-	M	N	K	L	F	L	I	L	L	L	I	F	S	H	E	V	F	S	-	-	-
nat_sp8	M	K	K	S	L	T	L	L	I	L	L	L	C	S	L	L	F	S	T	V	L	S
nat_sp9	M	K	K	L	L	P	I	L	I	G	L	S	L	S	G	F	S	S	L	S	Q	A

Sequence: MKKLLVIAALACGVATAQA

gen_sp	M	K	K	-	-	-	-	-	-	L	L	V	I	A	A	L	A	C	G	V	A	T	A	Q	A
nat_sp0	M	K	T	-	-	-	-	-	-	K	L	L	V	L	A	V	A	L	S	V	A	S	A	Q	A
nat_sp1	M	K	K	-	-	-	-	-	T	L	L	A	A	A	L	L	A	G	F	A	G	A	A	Q	A
nat_sp2	M	K	K	L	L	V	A	S	S	A	S	A	A	L	F	A	V	G	V	G	A	N	A	H	A
nat_sp3	M	K	K	-	-	-	-	T	A	I	A	I	A	V	A	L	A	G	F	A	T	V	A	Q	A
nat_sp4	M	K	K	-	-	-	-	T	A	I	A	I	T	V	A	L	A	G	F	A	T	V	A	Q	A
nat_sp5	M	K	K	L	A	T	L	T	A	L	A	G	A	L	T	M	A	-	V	A	T	A	A	Q	A
nat_sp6	M	K	F	-	-	-	-	-	F	V	L	V	A	I	A	F	A	L	L	A	C	V	A	Q	A
nat_sp7	M	K	R	-	-	-	-	K	V	L	A	L	V	I	P	A	L	L	A	A	G	A	A	H	A
nat_sp8	M	K	S	-	-	-	L	L	V	V	A	A	V	L	A	V	G	A	L	A	Q	G	D	D	A
nat_sp9	M	K	K	-	-	-	-	T	A	I	A	L	A	V	A	L	A	G	F	A	T	V	A	Q	A

Sequence: MKKRLHIGLLLSLIAFQAGFA

gen_sp	-	M	K	K	-	R	-	L	H	I	G	L	L	L	S	L	I	-	-	-	-	A	F	Q	A	G	F	A
nat_sp0	-	M	K	K	-	-	-	L	V	T	G	L	L	A	L	S	L	-	-	-	-	F	L	A	A	-	-	-
nat_sp1	-	M	K	K	-	R	A	L	A	I	A	F	L	L	A	S	L	-	-	-	-	I	P	S	A	A	Q	A
nat_sp2	-	M	K	R	-	-	-	L	V	T	G	L	L	A	L	S	L	-	-	-	-	F	L	A	A	-	-	-
nat_sp3	-	M	K	K	-	R	L	C	A	V	L	L	A	S	P	L	L	-	-	-	-	F	S	A	A	V	F	A
nat_sp4	-	M	I	K	-	R	T	L	T	V	S	L	L	S	L	S	L	G	A	M	F	A	S	A	G	V	M	A
nat_sp5	-	M	K	K	-	R	G	A	F	L	G	L	L	L	V	S	-	-	-	-	-	A	C	A	S	V	F	A
nat_sp6	-	M	I	R	-	-	-	L	L	I	A	L	F	L	L	L	A	-	-	-	-	S	V	A	P	G	F	A
nat_sp7	-	-	-	-	-	-	-	M	L	L	G	L	L	V	L	S	L	-	-	-	-	A	F	Q	G	T	L	A
nat_sp8	M	K	K	R	L	P	T	L	L	A	S	L	I	G	S	A	L	-	-	-	-	Y	S	Q	Q	A	L	A
nat_sp9	-	M	K	K	L	K	I	T	G	L	S	L	I	I	S	G	L	-	-	-	-	L	M	A	Q	A	Q	A

Sequence: MKKRVISALAALWLSVLGAPAVLA

gen_sp	-	M	K	K	R	V	I	S	A	L	A	-	A	L	W	L	S	V	L	G	A	P	A	V	L	A	A
nat_sp0	-	M	K	K	R	F	I	S	V	C	A	I	A	I	A	L	L	V	S	L	T	P	A	A	L	A	A
nat_sp1	-	M	M	K	S	F	F	S	A	A	A	-	-	L	L	L	G	L	V	A	P	S	A	V	L	A	A
nat_sp2	-	M	K	K	R	F	I	S	V	C	A	I	A	I	A	L	F	V	S	L	T	P	A	A	L	A	A
nat_sp3	-	M	T	R	Y	V	I	S	R	L	S	-	A	I	A	L	L	A	L	A	P	A	L	A	L	A	A
nat_sp4	-	M	K	K	L	L	I	S	A	V	S	-	-	-	-	A	L	V	L	G	S	G	A	A	L	A	A
nat_sp5	-	M	K	I	S	I	Y	A	T	L	A	-	-	-	A	L	S	L	A	L	P	A	V	A	-	-	A
nat_sp6	M	K	G	R	S	A	L	L	R	A	L	W	I	A	A	L	S	F	G	L	G	G	V	A	V	A	A
nat_sp7	-	M	K	K	V	V	N	S	V	L	A	-	-	S	A	L	A	L	T	V	A	P	M	A	F	A	A
nat_sp8	-	M	K	K	N	L	R	I	V	S	A	-	A	A	A	A	L	L	A	V	A	P	V	A	A	S	A
nat_sp9	-	M	K	K	L	T	V	A	I	S	A	V	A	A	S	V	L	M	A	M	S	A	Q	A	-	-	A

Sequence: MKKSLISFLALGLLFGSAFA

gen_sp	M	K	K	S	L	-	I	S	F	L	A	-	-	-	L	G	L	L	F	G	-	-	-	S	A	F	A
nat_sp0	M	K	K	F	L	K	S	T	A	A	L	A	L	G	L	S	L	T	F	G	L	F	S	P	A	Q	A
nat_sp1	-	M	K	S	L	-	L	P	I	S	S	-	-	-	L	L	V	L	L	G	-	S	A	S	A	F	A
nat_sp2	M	K	S	S	L	-	H	W	F	L	I	S	S	S	L	A	L	P	L	S	L	N	F	S	A	F	A
nat_sp3	M	K	K	R	L	-	L	G	I	A	L	-	-	-	G	S	L	L	F	T	-	T	S	S	A	V	A
nat_sp4	M	I	N	K	I	K	I	L	F	S	F	-	-	-	L	A	L	L	L	S	F	T	S	Y	A	K	A
nat_sp5	M	K	K	R	F	-	S	L	I	M	M	-	-	-	T	G	L	L	F	G	L	T	S	P	A	F	A
nat_sp6	M	K	K	L	L	-	I	S	A	V	-	-	-	-	S	A	L	V	L	G	-	S	G	A	A	F	A
nat_sp7	M	K	K	H	L	-	L	A	-	-	-	-	-	-	L	G	L	L	L	V	G	V	S	P	A	Q	A
nat_sp8	M	K	K	H	L	-	L	A	-	-	-	-	-	-	L	G	L	L	L	A	G	V	S	P	A	Q	A
nat_sp9	M	K	K	N	L	-	L	G	F	T	L	-	-	-	A	S	L	L	F	T	-	T	G	S	A	V	A

Sequence: MKKTGFIGKTLALVIAAGMAGTAFA

gen_sp	-	-	-	-	-	-	-	-	-	-	M	K	K	T	G	F	I	G	K	T	L	A	L	V	I	A	A	G	M	A	G	T	A	-	-	A	F	A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
nat_sp0	-	-	-	-	-	-	-	-	-	M	M	K	K	M	T	G	K	T	F	A	L	S	A	L	V	A	A	S	F	M	A	A	G	-	-	A	M	A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
nat_sp1	-	-	-	-	-	-	-	-	-	M	Q	K	K	R	I	G	K	S	V	V	A	A	L	A	I	I	A	M	S	A	G	T	V	A	A	W	A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
nat_sp2	-	-	-	-	-	-	-	-	-	-	M	T	T	I	V	K	R	A	L	V	A	A	G	M	V	L	A	I	G	G	A	-	-	A	Q	A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
nat_sp3	-	-	-	-	-	-	-	-	-	M	K	K	S	L	-	A	L	V	L	A	T	G	M	A	V	T	T	F	G	G	T	G	S	A	F	A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
nat_sp4	-	-	-	-	-	-	-	-	-	M	N	K	K	H	G	F	P	L	T	L	T	A	L	A	I	A	T	A	F	P	A	Y	-	-	A	A	Q	A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
nat_sp5	-	-	-	-	-	-	-	-	-	M	F	T	R	K	I	Q	K	T	A	L	-	A	M	L	I	S	G	A	M	A	G	T	-	-	A	Y	A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
nat_sp6	-	-	-	-	-	-	-	-	-	M	K	K	S	L	F	C	G	V	C	L	C	A	L	V	A	-	-	M	G	G	T	-	-	S	F	A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
nat_sp7	-	-	-	-	-	-	-	-	-	M	K	K	T	T	I	P	T	L	S	A	L	T	L	A	M	S	L	A	F	G	G	A	-	-	A	I	A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
nat_sp8	-	-	-	-	-	-	-	-	-	M	M	K	K	M	T	G	K	S	F	A	L	S	A	L	V	A	A	S	F	M	A	A	G	-	-	A	M	A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
nat_sp9	M	K	M	V	K	K	T	Y	G	F	R	K	S	K	I	S	K	T	L	C	G	A	V	L	G	T	V	A	A	V	S	V	A	G	Q	K	-	-	V	F	A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
					5				10						15					20																					25																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				

Sequence: MKLIPNKKTLIAGILAISTSFAYS

gen_sp	-	-	-	M	K	L	I	P	N	K	K	T	L	I	A	G	I	L	A	I	S	-	T	S	F	A	Y	S	-	-	-	-
nat_sp0	-	-	M	K	R	N	A	-	-	K	T	I	I	A	G	M	I	A	L	A	I	-	S	H	T	A	M	A	-	-	-	-
nat_sp1	-	-	-	M	K	N	V	-	-	K	T	L	I	A	A	A	I	L	S	S	M	-	S	F	A	S	F	A	-	-	-	-
nat_sp2	-	-	-	M	K	L	L	K	N	K	K	V	T	F	V	A	L	L	A	I	L	-	A	V	L	S	T	Q	S	V	S	A
nat_sp3	-	-	M	K	K	I	I	P	T	N	L	F	K	L	I	S	I	L	F	I	L	T	P	F	F	A	W	S	-	-	-	-
nat_sp4	-	-	-	M	K	N	W	-	-	K	T	L	L	L	G	I	A	M	I	A	N	-	T	S	F	A	-	-	-	-	-	-
nat_sp5	-	-	-	M	K	-	-	-	-	R	A	P	L	I	T	G	L	L	L	I	S	-	T	S	C	A	Y	A	-	-	-	-
nat_sp6	-	-	M	K	S	Q	K	I	G	S	M	I	L	L	I	G	I	L	L	A	I	-	F	N	F	A	Y	S	-	-	-	-
nat_sp7	-	-	-	M	K	K	L	P	N	K	S	L	I	A	L	A	L	L	S	V	S	G	A	S	F	G	H	G	Y	V	S	A
nat_sp8	M	Y	K	I	K	H	S	F	N	K	T	L	I	A	I	S	I	S	-	-	-	S	F	L	S	I	A	-	-	-	-	
nat_sp9	-	-	-	M	S	N	I	T	K	K	S	L	I	A	A	G	I	L	T	A	L	-	I	A	A	S	A	A	T	A	-	-

Sequence: MKLKKLGVLAICLGISSTFA

gen_sp	-	M	K	L	K	K	L	G	V	I	L	A	I	C	L	G	-	-	-	I	S	S	T	F	A	-	-	-	-	-	-	-
nat_sp0	-	M	K	R	P	L	D	F	L	L	A	I	C	L	I	L	L	-	-	R	S	S	T	F	A	-	-	-	-	-	-	-
nat_sp1	-	M	K	N	L	L	K	L	S	A	I	A	I	L	A	A	S	-	-	A	V	S	T	F	A	-	-	-	-	-	-	-
nat_sp2	M	Q	W	L	K	M	K	L	R	F	V	N	L	I	L	L	L	-	-	I	S	S	T	C	A	-	-	-	-	-	-	-
nat_sp3	-	M	A	L	K	S	K	F	L	V	G	S	I	L	A	T	F	-	-	I	L	N	G	F	S	-	-	S	P	A	Q	A
nat_sp4	M	K	K	L	V	S	S	V	I	L	A	L	I	L	F	G	F	S	W	V	S	P	A	F	A	-	-	-	-	-	-	-
nat_sp5	-	M	K	S	A	K	K	L	L	S	V	L	C	L	G	I	F	-	-	I	L	T	F	T	A	-	-	-	-	-	-	-
nat_sp6	-	M	K	L	S	S	L	P	S	G	L	G	L	A	S	L	L	G	L	I	S	S	A	T	A	-	-	-	-	-	-	-
nat_sp7	-	M	A	L	K	S	K	L	V	S	L	L	F	L	I	A	T	-	-	L	S	S	T	F	A	-	-	-	-	-	-	-
nat_sp8	-	-	M	L	K	K	L	G	V	L	L	S	A	I	V	L	V	-	-	I	A	S	F	F	V	T	V	T	P	A	L	A
nat_sp9	M	L	R	I	V	K	K	L	W	V	I	L	F	I	S	N	I	-	-	S	I	N	S	F	A	-	-	-	-	-	-	-

Sequence: MKLLTSFVLIGALAF

gen_sp	-	-	-	M	K	L	L	T	S	F	V	L	I	G	-	-	-	A	L	A	F	A	-
nat_sp0	-	-	M	R	A	L	L	F	T	S	V	V	L	L	A	L	-	A	F	V	E	A	-
nat_sp1	-	-	-	M	K	F	L	L	S	F	V	V	L	A	V	F	S	A	S	A	F	A	-
nat_sp2	-	-	M	T	P	L	L	T	L	F	L	V	A	L	I	G	L	P	L	A	Q	A	-
nat_sp3	-	-	M	R	L	P	L	L	S	F	V	I	F	A	L	F	-	A	L	V	F	A	-
nat_sp4	-	-	M	K	F	L	S	L	A	F	V	L	G	L	L	-	-	A	L	A	N	A	-
nat_sp5	M	L	L	R	S	L	T	S	A	F	V	L	S	A	-	-	-	G	L	A	Q	A	-
nat_sp6	-	-	M	K	I	L	L	F	V	T	L	I	A	L	A	F	V	A	L	C	S	A	-
nat_sp7	-	M	K	L	S	L	F	S	T	F	A	A	V	I	I	G	-	A	L	A	L	P	-
nat_sp8	-	M	K	K	V	L	L	L	L	F	V	L	T	I	-	-	-	G	L	A	L	S	A
nat_sp9	M	N	L	K	K	L	L	T	S	A	V	L	S	I	S	L	C	Q	S	A	F	A	-

Sequence: MKLSQSLTYLAVLGLAAGANA

gen_sp	-	M	K	L	S	Q	S	L	T	Y	L	A	V	L	G	L	A	A	G	A	N	A
nat_sp0	M	K	K	F	N	Q	S	L	L	A	T	A	M	L	L	A	A	G	G	A	N	A
nat_sp1	-	M	K	F	-	-	-	L	S	L	A	F	V	L	G	L	L	A	L	A	N	A
nat_sp2	M	K	K	L	L	S	T	L	L	L	S	L	G	L	-	A	A	G	L	A	Q	A
nat_sp3	M	K	S	L	S	-	-	L	L	L	A	V	A	L	G	L	A	T	A	V	S	A
nat_sp4	-	M	K	L	K	L	V	A	V	A	T	T	L	L	A	A	A	G	A	V	N	A
nat_sp5	-	M	K	L	S	Q	I	L	T	F	A	S	L	L	-	-	-	S	G	A	L	A
nat_sp6	-	M	L	T	S	L	S	L	T	A	L	A	L	L	-	-	-	P	S	A	N	A
nat_sp7	-	M	K	F	-	-	-	L	S	S	L	V	V	L	G	L	S	A	Q	A	L	A
nat_sp8	-	M	K	L	S	-	-	V	L	S	L	A	S	L	A	S	A	A	A	L	N	A
nat_sp9	M	K	R	L	S	-	-	L	A	M	V	T	L	L	-	A	C	A	G	A	Q	A

Sequence: MKMRTGKKGFLSILLAFLLVITSIPFTLVDVEA

gen_sp	M	K	M	R	T	G	K	K	G	F	L	S	I	-	L	L	A	F	L	L	V	I	T	S	I	P	F	-	T	L	V	D	V	E	A	
nat_sp0	M	K	M	R	T	G	K	K	G	F	L	S	I	-	L	L	A	F	L	L	V	I	T	S	I	P	F	-	T	L	V	D	V	E	A	
nat_sp1	-	-	M	S	R	S	L	K	K	F	V	S	I	-	L	L	A	A	A	L	L	I	P	I	G	R	-	-	L	A	P	V	A	E	A	
nat_sp2	M	N	Y	K	I	G	I	M	S	L	L	V	I	T	S	I	I	F	L	F	L	-	-	-	-	-	-	-	V	P	D	K	V	E	A	
nat_sp3	-	-	-	M	K	R	K	N	K	V	L	S	I	-	L	L	T	L	L	L	I	I	S	T	T	-	-	-	S	V	N	M	S	S	F	A
nat_sp4	-	-	M	K	S	I	T	K	I	F	L	I	L	G	L	F	A	F	L	L	V	A	F	-	-	-	-	-	A	P	S	S	S	V	A	
nat_sp5	-	-	-	-	-	M	T	K	P	F	S	V	-	-	L	L	A	S	L	L	V	I	T	S	T	-	-	-	V	S	P	L	A	S	A	
nat_sp6	-	M	M	K	K	L	I	Q	L	S	F	T	V	-	M	I	I	F	T	I	L	V	L	G	V	V	A	-	-	-	-	-	-	-	-	
nat_sp7	-	-	M	R	Q	K	A	I	F	K	I	A	V	-	L	L	A	F	I	G	L	S	L	M	V	S	S	I	Q	L	K	N	V	E	A	
nat_sp8	-	-	M	K	G	G	D	K	M	K	K	L	I	-	L	L	M	L	L	L	P	I	S	L	I	-	-	-	G	C	T	D	E	E	S	
nat_sp9	M	K	V	L	A	T	S	F	V	L	G	S	L	-	G	L	A	F	Y	L	P	L	V	V	-	-	-	-	T	T	P	K	T	L	A	
					5					10					15						20						25					30			35	

Sequence: MKNFATLSAVLAGATALA

gen_sp	-	-	M	K	N	F	A	T	-	-	-	-	-	-	-	-	-	L	S	A	V	L	A	G	A	T	A	L	A	
nat_sp0	-	-	-	M	A	F	A	K	L	S	A	-	-	-	-	-	-	F	V	L	A	L	G	A	T	V	A	L	G	
nat_sp1	-	-	-	M	K	F	S	S	A	L	V	-	-	-	-	-	-	L	S	A	V	A	A	T	-	-	A	L	A	
nat_sp2	-	-	M	T	R	F	L	I	-	-	-	-	-	-	-	-	-	L	S	A	V	L	A	G	P	-	A	L	A	
nat_sp3	-	-	-	M	K	F	A	-	-	-	-	-	-	-	-	-	-	L	L	A	L	A	A	V	-	-	A	L	A	
nat_sp4	M	Q	I	K	T	F	A	L	S	A	A	-	-	-	-	-	-	I	A	Q	V	A	T	L	-	-	A	L	A	
nat_sp5	-	-	-	M	R	F	T	A	T	V	L	S	R	V	A	T	G	L	A	L	G	L	S	M	A	T	A	S	L	A
nat_sp6	-	-	-	M	K	F	S	L	A	C	-	-	-	-	-	-	-	L	L	A	L	A	G	L	Q	A	A	L	A	
nat_sp7	M	K	K	T	A	I	A	L	A	V	A	-	-	-	-	-	-	L	A	G	F	A	T	V	-	-	A	Q	A	
nat_sp8	-	-	-	M	K	L	N	I	A	I	L	-	-	-	-	-	-	L	A	A	L	A	A	T	-	-	A	S	A	
nat_sp9	-	-	-	M	K	F	F	A	V	L	L	-	-	-	-	-	-	L	A	S	L	A	A	T	-	-	S	L	A	
					5					10					15						20						25			30

Sequence: MKVFTLAFAIICQLFASA

gen_sp	M	K	V	F	T	-	L	A	F	A	I	I	C	Q	L	F	A	S	A	-	-	-	-
nat_sp0	M	K	K	T	A	A	I	I	S	A	C	M	L	T	F	A	L	S	A	-	-	-	-
nat_sp1	M	K	T	Q	L	A	F	L	A	I	T	V	I	L	M	Q	L	F	A	Q	T	E	A
nat_sp2	M	K	Y	S	T	V	F	T	A	L	T	A	L	F	A	Q	A	S	A	-	-	-	-
nat_sp3	M	K	T	L	V	F	H	I	F	I	F	A	L	V	A	F	A	S	A	-	-	-	-
nat_sp4	M	K	Y	T	I	V	F	L	F	A	I	I	A	S	L	A	C	L	Q	L	T	F	A
nat_sp5	M	K	F	N	S	V	L	A	F	F	L	A	T	L	T	A	S	A	-	-	-	-	-
nat_sp6	M	K	L	S	V	V	L	A	L	F	I	I	F	Q	L	G	A	A	S	-	-	-	-
nat_sp7	M	K	F	L	I	A	F	V	A	I	A	F	F	A	C	V	S	A	-	-	-	-	-
nat_sp8	M	K	N	V	K	T	L	I	A	A	A	I	L	S	S	M	S	F	A	-	S	F	A
nat_sp9	M	K	F	S	T	-	L	S	T	V	A	A	I	A	A	F	A	S	A	-	-	-	-
					5					10					15						20		

Sequence: MKVFTLAFFLAIIIVSQA

gen_sp	-	-	M	K	V	F	T	L	-	-	-	-	-	A	F	F	L	A	I	I	V	S	Q	A	-
nat_sp0	-	-	-	-	M	A	K	V	L	F	S	L	F	S	F	L	F	L	I	I	G	V	S	A	-
nat_sp1	-	-	M	K	F	T	S	V	L	-	-	-	-	A	F	F	L	A	T	L	T	A	S	A	-
nat_sp2	-	-	M	K	K	A	V	I	L	F	S	L	F	C	F	L	C	A	I	P	V	V	Q	A	-
nat_sp3	M	K	K	T	V	F	R	L	N	F	L	T	A	C	I	S	L	G	I	V	S	Q	A	W	A
nat_sp4	-	-	M	K	A	T	-	-	-	-	-	-	-	L	I	F	F	L	L	A	Q	V	S	W	A
nat_sp5	-	-	M	K	F	I	I	R	T	I	L	I	A	L	F	L	I	A	I	I	N	E	S	Q	C
nat_sp6	-	-	M	F	N	T	R	L	A	I	-	-	-	F	L	L	L	I	V	V	S	L	S	Q	A
nat_sp7	M	Q	V	K	L	F	Y	T	-	-	-	-	-	L	A	L	W	A	P	I	L	V	S	A	-
nat_sp8	-	-	M	K	T	V	I	L	A	-	-	-	-	L	A	L	I	V	L	A	S	S	T	Q	A
nat_sp9	-	-	M	K	F	T	R	T	L	V	L	A	S	T	F	L	L	A	T	V	A	T	S	Q	A
					5				10					15					20					25	

Sequence: MLKKLAMAVGAMLTSSISFLLPSSAQA

gen_sp	-	-	M	L	K	K	L	A	M	A	V	G	A	M	L	T	S	I	S	F	L	L	-	P	S	S	A	Q	A	
nat_sp0	M	G	L	G	K	K	L	S	V	A	V	A	A	S	F	M	S	L	S	I	S	L	-	P	-	G	V	Q	A	
nat_sp1	M	G	L	G	K	K	L	S	V	A	V	A	A	S	F	M	S	L	T	I	S	L	-	P	-	G	V	Q	A	
nat_sp2	-	-	-	M	K	K	S	L	L	A	V	A	V	A	G	A	V	L	L	S	S	-	-	-	-	-	A	V	Q	A
nat_sp3	-	-	-	M	L	K	K	L	I	T	T	A	V	L	A	M	L	I	F	-	-	-	-	-	-	-	T	L	A	A
nat_sp4	-	-	M	K	L	K	L	I	L	A	V	A	M	L	A	F	S	L	-	-	-	-	-	-	-	-	P	S	Q	A
nat_sp5	-	-	-	-	M	K	K	L	A	I	M	A	A	A	S	M	I	F	T	V	G	-	-	-	-	-	S	A	Q	A
nat_sp6	-	-	-	M	K	K	T	T	L	A	M	S	A	L	A	L	S	L	G	L	A	L	S	P	L	S	A	T	A	
nat_sp7	-	-	-	M	L	K	K	S	L	A	A	L	A	L	G	T	A	L	L	S	A	G	-	-	-	-	Q	A	M	A
nat_sp8	-	-	M	N	L	K	K	L	L	T	S	A	V	L	S	I	S	L	C	Q	-	-	-	-	-	-	S	A	F	A
nat_sp9	-	-	-	-	M	K	K	I	A	I	V	G	A	L	L	T	S	F	V	A	S	-	-	-	-	-	S	V	W	A
					5				10					15					20					25						

Sequence: MLKRFTLFLGLALASSLA

gen_sp	-	M	L	K	R	F	-	L	T	L	F	L	-	-	-	G	F	L	A	L	A	S	S	L	A	-	-	-	-	-	-	
nat_sp0	-	-	-	M	R	F	L	T	V	A	F	-	-	-	-	-	-	L	F	L	A	L	S	A	S	A	L	A	-	-	-	
nat_sp1	-	-	M	K	L	F	Q	L	T	L	F	L	L	V	N	A	F	L	A	L	A	S	S	-	-	-	-	-	-	-	-	
nat_sp2	-	-	M	R	L	Y	F	R	K	L	W	L	-	-	-	T	N	L	F	L	G	G	A	L	A	S	S	A	-	-	-	
nat_sp3	M	A	L	K	T	L	Q	A	L	I	F	L	-	-	-	G	-	L	F	A	A	S	C	L	A	-	-	-	-	-	-	
nat_sp4	-	M	L	K	K	T	L	L	G	L	T	A	-	-	-	G	A	L	L	L	N	A	S	S	A	L	A	-	-	-	-	
nat_sp5	-	-	M	K	R	F	A	L	S	L	L	A	-	-	-	G	L	V	A	L	Q	A	S	A	-	-	-	-	-	-	-	
nat_sp6	-	-	M	F	K	F	A	L	T	L	T	-	-	-	-	-	-	L	C	L	A	G	S	L	S	L	A	-	-	-	-	
nat_sp7	-	-	M	F	K	F	A	L	A	L	T	-	-	-	-	-	-	L	C	L	A	G	A	S	L	S	L	A	-	-	-	
nat_sp8	-	-	M	K	K	L	T	L	V	L	F	-	-	-	-	G	M	L	F	L	A	S	S	A	H	A	-	-	-	-	-	
nat_sp9	-	M	L	S	L	R	S	L	L	P	H	L	-	-	-	G	-	L	F	L	C	L	A	L	H	L	S	P	S	L	S	A
					5				10					15					20					25				30				

Sequence: MLKRFVKLAVIALAFAYVSA

gen_sp	-	M	L	K	R	F	V	K	L	A	V	I	A	L	A	F	A	-	Y	V	S	A
nat_sp0	-	M	R	I	K	F	L	V	V	L	A	V	I	C	L	F	A	H	Y	A	S	A
nat_sp1	-	-	-	M	K	F	S	Q	A	V	I	A	L	A	A	A	T	-	V	V	S	A
nat_sp2	M	L	K	M	N	V	K	K	A	L	V	I	L	V	A	L	A	-	L	V	A	A
nat_sp3	-	M	K	Y	K	K	L	S	V	A	V	A	A	F	A	F	A	-	A	V	S	A
nat_sp4	-	M	K	Y	N	K	L	S	V	A	V	A	A	F	A	F	A	-	A	V	S	A
nat_sp5	-	-	-	M	Q	L	T	Q	V	L	A	V	A	I	L	A	A	-	G	V	S	A
nat_sp6	-	-	-	-	M	K	F	I	V	L	V	L	F	C	A	V	A	-	Y	V	S	A
nat_sp7	-	-	-	M	K	F	L	I	A	F	V	A	I	A	F	F	A	-	C	V	S	A
nat_sp8	-	-	M	A	R	F	T	I	V	L	A	V	L	F	A	A	A	-	L	V	S	A
nat_sp9	-	M	R	I	K	F	L	V	V	L	A	V	I	C	L	L	A	H	Y	A	S	A

Sequence: MNIRLGALLAGLLLSAMASAVFA

gen_sp	M	N	I	R	L	G	A	L	L	-	-	-	-	A	G	L	L	L	-	-	S	A	M	A	S	A	V	F	A					
nat_sp0	-	M	N	K	R	G	A	L	-	-	-	-	-	L	S	L	L	L	-	-	L	S	A	S	V	S	A	F	A					
nat_sp1	-	M	K	K	R	G	A	F	-	-	-	-	-	L	G	L	L	L	-	-	V	S	A	C	A	S	V	F	A					
nat_sp2	-	M	N	K	R	G	A	L	-	-	-	-	-	L	S	L	L	F	-	-	L	S	A	S	V	S	A	F	A					
nat_sp3	-	M	K	K	R	G	A	F	-	-	-	-	-	L	G	R	L	L	-	-	V	S	A	C	A	S	V	F	A					
nat_sp4	-	M	R	S	L	G	A	L	-	-	-	-	-	-	L	L	L	L	-	-	L	S	A	C	L	A	V	S	A					
nat_sp5	-	M	A	R	L	G	A	L	L	L	A	A	A	L	G	A	L	L	S	F	A	L	L	A	A	A	V	A	S					
nat_sp6	-	M	K	K	R	L	C	A	V	L	L	-	-	A	S	P	L	L	-	-	-	-	F	S	A	A	V	F	A					
nat_sp7	-	M	K	T	R	L	A	L	P	C	L	-	-	L	G	S	L	L	-	-	-	-	L	S	S	A	V	H	A					
nat_sp8	-	M	I	R	K	G	A	A	L	-	-	-	-	A	G	L	V	L	-	-	-	-	M	S	P	V	I	A						
nat_sp9	-	M	Q	R	L	G	A	T	L	-	-	-	-	L	C	L	L	L	-	-	-	A	A	A	V	P	T	A	P					
										5					10					15					20					25				

Sequence: MNKKFKTIMALAIATLSAAGVGVAHA

gen_sp	M	N	K	K	F	K	T	I	M	A	L	A	I	A	T	-	L	S	A	A	G	V	G	-	-	-	V	A	H	A
nat_sp0	-	-	-	M	K	K	T	L	I	A	L	A	I	A	A	-	-	-	S	A	A	S	G	-	-	-	M	A	H	A
nat_sp1	-	M	K	K	T	I	M	A	S	S	L	A	V	A	L	G	V	T	G	Y	A	A	G	T	G	H	Q	A	H	A
nat_sp2	M	N	K	P	T	K	L	F	S	T	L	A	L	A	A	G	M	T	A	A	A	A	G	G	A	G	T	I	H	A
nat_sp3	M	N	F	K	K	T	V	V	S	A	L	S	I	S	A	-	L	A	L	S	V	S	G	-	-	-	V	A	S	A
nat_sp4	-	-	-	M	K	K	T	L	I	A	L	A	V	A	A	-	-	S	A	A	V	S	G	-	-	-	S	V	M	A
nat_sp5	M	N	M	K	K	F	V	K	K	P	L	A	I	A	V	-	L	M	L	A	S	G	G	M	V	N	M	V	H	A
nat_sp6	-	M	K	L	K	K	T	I	G	A	M	A	L	A	T	-	-	L	F	A	T	M	G	-	-	-	-	A	S	A
nat_sp7	-	M	N	K	K	W	L	N	I	P	A	L	I	A	L	-	L	A	A	I	A	F	G	S	V	A	P	A	E	A
nat_sp8	-	M	K	K	T	I	M	A	S	S	L	A	V	A	L	G	V	T	G	Y	A	A	S	T	G	H	E	A	H	A
nat_sp9	-	-	-	-	M	K	K	G	K	I	L	A	L	A	G	-	V	A	L	L	A	T	G	-	-	-	V	L	A	A

Sequence: MQKKTAIAIAAGTAIATVAAGTQA

gen_sp	M	Q	K	K	T	A	I	A	-	I	A	A	G	T	A	I	A	T	V	A	A	G	-	-	-	-	T	Q	A
nat_sp0	-	M	Q	K	K	V	I	A	A	I	I	G	T	S	A	I	S	A	V	A	A	T	-	-	-	Q	A	N	A
nat_sp1	-	M	K	K	T	A	I	A	I	A	V	A	L	A	G	F	A	T	V	A	Q	A	-	-	-	-	-	-	-
nat_sp2	-	M	K	K	T	A	I	A	I	T	V	A	L	A	G	F	A	T	V	A	Q	A	-	-	-	-	-	-	-
nat_sp3	-	M	K	K	T	A	I	A	L	A	V	A	L	A	G	F	A	T	V	A	Q	A	-	-	-	-	-	-	-
nat_sp4	-	M	K	K	T	A	I	A	L	A	V	A	L	V	G	F	A	T	V	A	Q	A	-	-	-	-	-	-	-
nat_sp5	-	M	K	K	T	A	I	A	L	V	V	A	G	L	A	A	A	S	V	A	Q	A	-	-	-	-	-	-	-
nat_sp6	-	M	K	K	I	A	T	A	-	T	I	A	T	A	G	I	A	T	F	A	F	A	H	-	H	D	A	Q	A
nat_sp7	-	M	Q	K	K	Y	I	T	A	I	I	G	T	T	A	L	S	A	L	A	S	T	-	-	-	H	A	Q	A
nat_sp8	-	M	K	K	T	L	L	A	T	A	I	A	G	A	M	A	A	S	G	A	Q	A	-	-	-	-	-	-	-
nat_sp9	-	M	K	K	I	A	T	A	-	T	I	A	T	A	G	F	A	T	I	A	I	A	S	G	N	Q	A	H	A

Sequence: MRLIVFLATSATSLFASLA

gen_sp	-	-	M	R	L	I	V	F	L	A	T	S	A	T	S	L	F	A	S	L	A	-	-	-	-	-	-
nat_sp0	-	M	Y	S	L	I	K	S	L	A	T	F	A	T	L	F	S	L	T	L	A	-	-	-	-	-	-
nat_sp1	-	-	M	R	F	L	T	V	A	F	L	F	L	A	L	S	A	S	A	L	A	-	-	-	-	-	-
nat_sp2	-	-	M	R	F	T	L	I	E	A	V	A	L	T	A	V	S	L	A	S	A	-	-	-	-	-	-
nat_sp3	-	-	M	R	L	I	V	L	S	L	L	F	T	S	T	L	A	-	-	-	-	-	-	-	-	-	-
nat_sp4	-	-	M	Q	V	I	V	L	P	L	V	F	L	A	T	F	A	T	S	G	S	L	A	-	-	-	-
nat_sp5	M	L	R	L	A	I	P	L	F	L	F	A	L	C	S	F	T	L	F	S	S	A	-	-	-	-	
nat_sp6	M	V	K	S	V	L	A	S	A	L	F	A	V	S	A	L	A	-	-	-	-	-	-	-	-	-	
nat_sp7	M	H	R	I	F	L	I	T	V	A	L	A	L	L	T	A	S	P	A	S	A	-	-	-	-	-	
nat_sp8	-	-	M	R	S	L	H	I	L	L	V	F	T	A	S	L	L	A	S	L	T	E	S	A	K	A	
nat_sp9	M	K	R	K	T	L	P	L	L	A	L	V	A	T	S	L	F	L	S	A	-	-	-	-	-	-	
510152025																											

Sequence: MRRLFLSSLASLSVASA

gen_sp	-	-	-	M	R	R	L	F	L	L	S	S	L	A	S	L	S	V	A	S	A	-	-	-
nat_sp0	-	M	R	K	L	T	L	T	L	S	A	L	A	L	A	L	S	L	N	S	V	A	D	A
nat_sp1	-	-	-	-	M	R	S	F	L	L	A	S	L	A	S	L	S	V	I	S	V	Y	G	-
nat_sp2	M	R	R	L	I	P	I	L	L	G	S	L	V	L	S	L	S	I	L	V	A	P	A	-
nat_sp3	-	M	S	I	R	R	L	A	C	S	L	L	L	S	S	L	A	L	P	V	L	A	A	-
nat_sp4	-	M	R	N	R	L	L	S	L	V	T	L	F	L	S	L	S	V	A	T	A	V	S	A
nat_sp5	-	-	-	M	R	Q	L	L	L	T	L	S	L	I	S	V	S	A	S	D	A	-	-	-
nat_sp6	M	R	L	R	E	F	T	A	L	S	S	L	L	F	S	L	L	L	L	S	A	S	A	-
nat_sp7	-	-	-	M	R	S	L	V	L	L	S	S	V	L	A	L	V	A	P	S	K	G	-	-
nat_sp8	-	-	-	M	R	G	L	T	L	L	S	L	A	F	-	L	G	V	C	S	A	-	-	-
nat_sp9	-	-	-	-	M	R	I	F	L	L	C	L	A	L	S	L	S	V	F	A	-	-	-	-

Sequence: MSNKPAKCLAVLAAIATLSATQA

gen_sp	-	-	M	S	N	K	P	-	A	K	C	L	-	-	-	A	V	L	A	A	I	A	T	L	S	A	T	Q	A	-	-	-			
nat_sp0	M	N	K	T	W	K	K	-	A	A	T	V	-	-	-	L	A	F	A	G	I	A	-	L	S	A	T	A	-	-	-	-			
nat_sp1	-	M	K	N	W	I	K	-	-	-	-	-	-	-	-	V	A	V	A	A	I	A	-	L	S	A	A	T	V	Q	A	-			
nat_sp2	-	-	M	K	W	C	K	R	G	Y	V	L	-	-	-	A	A	I	L	A	L	A	-	S	A	T	I	Q	A	-	-	-			
nat_sp3	-	-	-	-	M	N	K	P	S	K	F	A	-	-	-	L	A	L	A	F	A	A	-	V	T	A	S	G	V	A	S	A			
nat_sp4	-	-	M	I	S	N	K	-	I	A	I	L	-	-	-	L	A	V	L	V	V	A	-	V	A	C	A	Q	A	-	-	-			
nat_sp5	-	-	-	-	-	-	-	-	M	K	A	K	-	-	-	L	L	V	L	L	C	A	-	L	S	A	T	D	A	-	-	-			
nat_sp6	-	-	M	A	Y	S	K	-	I	T	L	L	-	-	-	A	A	L	A	A	I	A	-	Y	A	Q	T	Q	A	-	-	-			
nat_sp7	-	M	S	F	F	T	K	-	T	A	Q	L	V	S	G	A	A	V	A	A	T	L	-	F	T	A	T	A	Q	A	-	-			
nat_sp8	-	-	-	-	M	S	P	-	T	A	C	V	-	-	-	L	V	L	A	L	A	A	-	L	R	A	T	G	-	-	-	-			
nat_sp9	-	M	P	S	A	K	P	-	L	F	C	L	-	-	-	A	T	L	A	G	A	A	-	L	A	A	P	-	-	-	-	-			
					5					10											15									20			25		30

Sequence: MTKFLLSLIFITIASALA

gen_sp	-	M	T	K	F	L	L	S	L	I	F	-	-	I	T	I	A	S	A	L	A	-	-	
nat_sp0	M	T	K	P	L	L	L	S	L	I	F	-	-	F	I	I	P	A	A	L	A	-	-	
nat_sp1	M	A	K	P	L	W	L	S	L	I	L	-	-	F	I	I	P	V	A	L	A	-	-	
nat_sp2	-	M	K	L	F	S	I	F	L	I	F	T	I	F	I	I	A	S	A	L	V	A	A	
nat_sp3	M	M	T	K	I	K	L	L	M	L	I	I	I	F	Y	L	I	I	S	A	S	A	H	A
nat_sp4	-	M	L	Q	K	F	L	I	C	L	S	L	I	F	T	L	A	S	A	-	-	-	-	
nat_sp5	-	-	-	M	K	L	L	L	V	L	L	-	-	T	I	A	S	V	A	L	A	-	-	
nat_sp6	-	-	-	M	K	L	L	L	V	L	I	-	-	T	L	I	I	A	A	L	A	-	-	
nat_sp7	M	T	K	H	A	R	F	F	L	L	P	S	F	I	L	I	S	A	A	L	I	A	G	
nat_sp8	-	-	-	M	K	F	L	S	V	L	S	L	A	I	T	L	A	A	A	-	-	-	-	
nat_sp9	-	M	K	Y	S	L	I	F	I	L	T	L	A	C	L	I	A	S	S	L	A	-	-	
					5					10										15			20	

Sequence: MTKLLAVIAASLMFAASTFA

gen_sp	-	-	-	-	M	T	K	L	L	A	V	I	A	A	S	L	M	F	A	A	S	T	-	-	-	-	F	A
nat_sp0	-	-	-	M	R	K	S	L	L	A	I	L	A	V	S	S	L	V	F	S	S	A	-	-	-	S	F	A
nat_sp1	-	-	-	-	-	M	K	L	L	A	I	I	S	A	S	L	A	L	A	G	F	T	-	-	-	-	T	A
nat_sp2	-	-	-	M	K	K	Q	L	L	S	A	L	I	G	A	S	L	L	A	P	M	A	-	-	-	A	S	A
nat_sp3	M	K	T	T	V	T	K	L	L	A	T	V	A	A	A	S	T	I	F	G	M	S	T	L	P	A	F	A
nat_sp4	-	-	M	R	T	L	K	S	L	V	I	V	S	A	A	L	L	P	F	S	A	T	-	-	-	A	F	A
nat_sp5	-	-	-	-	-	-	M	K	L	S	L	S	V	A	L	S	L	A	A	S	T	-	-	-	A	Q	A	
nat_sp6	-	-	-	-	M	K	K	T	L	L	I	A	A	S	L	S	F	F	S	A	S	-	-	-	A	L	A	
nat_sp7	-	-	-	-	M	K	R	L	S	I	A	I	T	S	L	L	M	A	A	S	A	S	-	-	-	T	I	A
nat_sp8	-	-	-	-	M	K	K	T	L	L	A	V	A	I	G	G	A	M	F	A	T	S	-	-	-	A	A	A
nat_sp9	-	-	-	-	M	K	N	L	L	K	L	S	A	I	A	I	L	A	A	S	A	V	S	-	-	T	F	A
					5					10					15					20							25	

Sequence: MTRSLFIFSL LALAIFSGVSASA

gen_sp	-	M	T	R	S	L	F	I	F	S	L	L	A	L	A	I	F	S	G	V	S	A	S	A	-	-
nat_sp0	M	I	T	R	W	L	L	I	T	S	F	L	A	L	A	I	L	S	L	S	S	A	-	-	-	-
nat_sp1	M	A	F	R	V	L	L	L	F	S	L	T	A	L	L	I	F	S	A	V	S	P	S	F	A	-
nat_sp2	-	M	K	T	S	I	L	F	V	I	F	S	L	A	L	L	F	A	L	S	A	A	-	-	-	-
nat_sp3	-	-	-	-	M	R	L	S	V	S	L	L	A	L	A	F	G	S	L	V	A	A	-	-	-	-
nat_sp4	-	M	K	K	R	F	M	M	F	T	L	L	A	A	V	F	S	G	V	A	H	A	-	-	-	-
nat_sp5	-	M	T	R	Y	F	S	I	V	L	S	L	L	L	A	V	S	C	V	F	L	P	V	A	S	A
nat_sp6	M	G	I	T	R	N	L	I	L	G	L	A	C	L	A	F	V	S	I	A	K	A	-	-	-	-
nat_sp7	-	M	T	K	P	F	S	V	V	L	A	S	L	L	A	I	T	S	T	V	S	P	L	A	S	A
nat_sp8	-	M	T	K	P	F	S	V	L	L	A	S	L	L	V	I	T	S	T	V	S	P	L	A	S	A
nat_sp9	M	K	Q	R	S	F	L	S	I	L	C	F	I	L	L	A	F	G	V	A	S	V	S	A	-	-

Sequence: MTSYEFLLVILGVLLSGA

gen_sp	-	-	M	T	S	Y	E	F	L	L	V	I	L	G	V	L	L	S	G	A	-	-	-	-	-	
nat_sp0	M	T	T	S	S	Y	-	F	L	L	V	A	L	G	L	L	L	Y	V	F	Q	S	S	F	G	
nat_sp1	M	T	T	S	S	Y	-	F	L	L	V	A	L	G	L	L	L	Y	V	C	R	S	S	F	G	
nat_sp2	M	T	T	S	S	Y	-	F	L	L	V	A	L	G	L	L	L	Y	L	C	Q	S	S	F	G	
nat_sp3	M	K	T	S	V	F	-	V	L	V	L	G	L	V	L	L	F	A	V	S	F	A	-	-	-	
nat_sp4	M	T	T	S	S	Y	-	F	L	L	V	A	L	G	L	L	L	Y	V	C	Q	S	S	F	G	
nat_sp5	M	T	T	S	S	Y	-	F	L	L	V	T	L	G	L	L	L	Y	V	C	R	S	S	F	G	
nat_sp6	M	T	T	S	F	Y	-	F	L	L	V	A	L	G	L	L	L	Y	V	C	Q	S	-	-	-	
nat_sp7	M	T	T	S	S	Y	-	F	L	L	V	A	L	G	L	L	L	Y	V	R	Q	S	-	F	S	
nat_sp8	M	T	T	S	S	Y	-	F	L	L	V	A	L	G	L	L	L	Y	V	C	Q	S	-	-	-	
nat_sp9	M	A	T	S	S	C	-	F	L	L	V	T	L	G	L	L	L	H	V	Q	Q	A	-	-	-	
		5					10					15					20					25				

Sequence: MVSFKSALFAAAAVATVADA

gen_sp	M	V	S	F	-	-	-	K	S	A	L	F	A	A	A	A	V	A	T	V	A	D	A	-	-	-	-	-	
nat_sp0	-	M	K	F	-	-	-	G	S	A	L	V	A	A	V	A	A	V	A	G	V	A	A	-	-	-	-	-	
nat_sp1	M	V	S	K	-	-	-	S	L	F	L	A	A	A	V	N	L	A	G	V	L	A	-	-	-	-	-	-	
nat_sp2	M	L	S	F	-	-	-	K	S	L	L	A	A	A	V	V	A	S	S	A	L	A	-	-	-	-	-	-	
nat_sp3	M	L	S	F	V	K	K	S	I	A	L	V	A	A	L	Q	A	V	T	A	L	A	-	-	-	-	-	-	
nat_sp4	M	F	K	F	-	-	-	S	A	S	L	A	A	A	L	A	V	P	F	V	A	A	-	-	-	-	-	-	
nat_sp5	M	V	S	L	-	-	-	K	S	V	L	A	A	A	T	A	V	S	S	A	I	A	-	-	-	-	-	-	
nat_sp6	M	R	V	S	-	-	-	A	F	A	L	L	A	A	A	T	A	A	A	-	-	-	-	-	-	-	-		
nat_sp7	M	V	S	F	-	-	-	K	S	L	L	V	A	V	S	A	L	T	G	A	L	A	-	-	-	-	-		
nat_sp8	-	M	S	F	K	T	L	S	A	L	A	L	A	L	G	A	A	V	Q	F	A	S	A	-	-	-	-	-	
nat_sp9	-	M	S	F	I	-	-	R	S	A	L	A	A	A	A	F	V	A	L	S	I	G	A	V	Q	T	A	S	A
					5			10						15					20					25					

Sequence: MVSFSSLLAAASLAVVNA

gen_sp	M	V	S	F	S	S	L	L	A	A	A	S	L	A	V	-	V	N	A
nat_sp0	M	V	S	F	S	S	L	V	L	A	A	S	T	V	A	G	V	L	A
nat_sp1	M	V	S	F	S	S	L	L	L	A	C	S	A	V	T	-	A	F	A
nat_sp2	M	V	S	K	S	L	F	L	A	A	A	V	N	L	A	G	V	L	A
nat_sp3	M	V	S	F	S	S	L	L	L	A	V	S	A	V	S	G	A	L	A
nat_sp4	M	V	S	F	S	S	L	V	L	A	L	S	T	V	A	G	V	L	A
nat_sp5	M	V	S	F	S	S	L	A	L	A	L	S	T	V	V	G	V	L	A
nat_sp6	M	V	S	F	S	S	I	L	L	A	C	S	A	A	I	G	A	L	A
nat_sp7	M	V	S	F	T	S	L	L	A	G	V	A	A	I	S	G	V	L	A
nat_sp8	M	R	V	S	S	S	L	I	A	L	A	A	L	A	V	Q	A	L	A
nat_sp9	M	V	S	F	T	Y	L	L	A	A	V	S	A	V	T	G	A	V	A

Sequence: MVSFSSLNALFLATVLA

gen_sp	M	V	S	F	S	S	L	N	A	-	-	-	-	L	F	L	-	-	-	A	T	V	L	A
nat_sp0	M	V	S	F	S	S	L	A	-	-	-	-	-	L	A	L	S	T	V	V	G	V	L	A
nat_sp1	M	V	S	F	S	S	C	L	-	-	-	-	-	R	A	L	A	L	G	S	S	V	L	A
nat_sp2	M	V	S	F	S	S	L	V	-	-	-	-	-	L	A	L	S	T	V	A	G	V	L	A
nat_sp3	M	V	S	L	K	S	L	A	A	I	L	V	A	M	F	L	A	T	G	P	T	V	L	A
nat_sp4	M	R	V	S	S	S	L	I	-	-	-	-	-	A	L	A	A	L	A	V	Q	A	L	A
nat_sp5	M	V	S	F	S	S	L	V	-	-	-	-	-	L	A	A	S	T	V	A	G	V	L	A
nat_sp6	-	M	S	F	S	S	L	R	R	A	L	-	-	V	F	L	G	A	C	S	S	A	L	A
nat_sp7	M	V	S	F	S	S	L	L	-	-	-	-	-	L	A	V	S	A	V	S	G	A	L	A
nat_sp8	M	V	A	F	S	S	L	I	-	-	-	-	-	C	A	L	T	S	I	A	S	T	L	A
nat_sp9	-	M	A	V	S	S	T	P	W	A	L	V	A	L	F	L	M	A	S	S	T	V	M	A
					5					10					15					20				

Sequence: MVSNKRVLALSALFGCCSLASA

gen_sp	-	M	V	S	N	K	R	V	L	A	L	S	A	L	F	G	-	-	-	C	C	S	L	A	S	A	-
nat_sp0	-	-	-	-	-	M	K	V	L	A	L	S	A	L	L	-	-	-	-	-	-	S	L	A	S	A	-
nat_sp1	-	-	-	-	-	M	K	V	L	A	L	S	A	L	L	-	-	-	-	-	-	S	L	A	S	A	A
nat_sp2	-	-	-	M	R	V	S	V	P	V	L	A	L	A	F	G	-	-	-	-	-	S	L	A	A	A	-
nat_sp3	M	V	S	P	R	V	R	L	A	A	L	A	L	S	V	C	A	I	L	C	L	G	M	H	A	S	A
nat_sp4	M	A	I	S	K	I	A	F	L	A	L	I	A	L	S	G	-	-	-	L	C	G	L	A	S	A	-
nat_sp5	-	-	-	-	M	V	K	S	V	L	A	S	A	L	F	A	-	-	-	V	S	A	L	A	-	-	-
nat_sp6	M	K	S	R	Y	K	R	L	T	S	L	A	L	S	L	S	M	A	L	G	I	S	L	P	A	W	A
nat_sp7	-	-	M	N	K	T	L	I	L	A	L	S	A	L	L	-	-	-	-	-	-	G	L	A	A	-	-
nat_sp8	M	V	W	A	N	A	K	M	R	L	A	L	S	L	V	T	V	F	F	G	I	S	L	A	-	-	-
nat_sp9	-	-	-	M	V	S	P	V	L	A	L	F	S	A	F	-	-	-	-	L	C	H	V	A	I	A	-

Sequence: MVTMKLRLIALAVCLCTFINASFA

gen_sp	-	-	-	-	-	M	V	T	M	K	L	R	L	I	A	L	A	V	C	L	C	T	F	I	N	A	S	F	A	-
nat_sp0	-	-	-	-	-	-	-	-	M	K	F	R	L	T	A	L	A	V	A	A	L	L	T	S	T	A	S	F	A	-
nat_sp1	-	-	-	-	-	M	V	T	M	K	L	F	A	L	V	L	L	V	S	A	S	V	F	A	-	-	-	-	-	-
nat_sp2	-	-	-	-	-	M	A	K	L	I	P	T	I	A	L	V	S	V	L	L	F	I	I	A	N	A	S	F	A	-
nat_sp3	-	-	-	-	-	M	D	V	T	R	L	L	L	A	T	L	L	V	C	L	C	F	F	T	A	S	S	-	-	-
nat_sp4	-	-	-	-	-	M	M	T	M	L	R	G	W	I	T	M	L	V	M	L	T	A	I	N	A	Q	A	-	-	-
nat_sp5	-	-	-	-	-	-	-	-	M	K	L	L	K	A	L	A	V	L	S	L	A	T	I	S	S	H	S	F	A	-
nat_sp6	M	K	T	P	R	L	K	K	L	A	L	V	C	A	L	G	F	A	C	I	T	F	S	A	I	N	A	V	Q	A
nat_sp7	-	-	-	-	-	M	I	R	M	S	K	R	L	G	V	I	L	F	V	S	C	-	I	S	I	N	S	F	A	-
nat_sp8	-	-	-	-	-	-	-	-	-	-	-	-	M	K	L	L	I	L	A	L	C	F	A	A	A	S	A	-	-	-
nat_sp9	M	N	T	I	F	S	A	R	I	M	K	R	L	A	L	T	T	A	L	C	T	A	F	I	S	A	A	H	A	-
						5						10																	30	

Sequence: MYSLIPSLAVLAALSFAVSA

gen_sp	-	-	M	Y	S	L	I	P	S	-	-	L	A	V	L	A	A	L	S	F	A	V	S	A	-
nat_sp0	-	-	M	S	K	S	L	A	G	-	-	L	A	V	L	A	A	L	F	I	A	V	D	A	-
nat_sp1	-	-	-	-	M	S	I	R	L	-	-	I	A	V	L	S	A	A	S	I	A	V	T	S	A
nat_sp2	-	-	M	K	Q	T	I	C	G	-	-	L	A	V	L	A	A	L	S	S	A	P	V	F	A
nat_sp3	-	-	M	K	S	L	L	P	I	-	-	S	S	L	L	V	L	L	G	S	A	S	V	S	A
nat_sp4	-	-	M	K	K	S	L	I	A	-	-	L	A	V	L	A	A	S	G	A	A	M	A	-	-
nat_sp5	-	-	M	K	K	L	L	P	L	A	V	L	A	A	L	S	S	V	H	V	A	S	A	Q	A
nat_sp6	-	-	M	K	K	I	I	S	L	-	-	A	L	A	L	A	L	S	A	S	A	-	-	-	-
nat_sp7	M	A	I	S	K	L	I	P	T	-	-	L	V	L	F	V	L	F	S	F	D	V	S	V	A
nat_sp8	-	-	M	K	S	L	L	P	-	-	-	L	A	I	L	A	A	L	A	V	A	A	L	C	-
nat_sp9	-	-	M	S	L	R	S	V	L	-	-	V	A	A	L	A	A	L	A	V	A	-	-	-	-
						5						10													25

Supp Section VI: Bibliography

- (1) Kingma, D. P.; Welling, M. Auto-Encoding Variational Bayes. In *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*; 2014.
- (2) Madeira, F.; Park, Y. M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A. R. N.; Potter, S. C.; Finn, R. D.; Lopez, R. The EMBL-EBI Search and Sequence Analysis Tools APIs in 2019. *Nucleic Acids Res.* **2019**. <https://doi.org/10.1093/nar/gkz268>.
- (3) Eddy, S. R. Profile Hidden Markov Models. *Bioinformatics.* 1998. <https://doi.org/10.1093/bioinformatics/14.9.755>.
- (4) Owji, H.; Nezafat, N.; Negahdaripour, M.; Hajiebrahimi, A.; Ghasemi, Y. A Comprehensive Review of Signal Peptides: Structure, Roles, and Applications. *Eur. J. Cell Biol.* **2018**, 97 (6), 422–441. <https://doi.org/10.1016/j.ejcb.2018.06.003>.
- (5) Almagro Armenteros, J. J.; Tsirigos, K. D.; Sønderby, C. K.; Petersen, T. N.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 5.0 Improves Signal Peptide Predictions Using Deep Neural Networks. *Nat. Biotechnol.* **2019**, 37 (4), 420–423. <https://doi.org/10.1038/s41587-019-0036-z>.